

COUNTERFEIT PROTECTION USING DNA

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The following commonly-owned US Provisional Patent Application Nos. 63/582,199 and 63/623,085 contain subject matter related to that described herein, and each of these applications is hereby incorporated by reference in its entirety to the fullest extent permitted by applicable law.

FIELD OF THE INVENTION

[0002] The invention relates generally to the field of synthetic biology, and more specifically to methods of counterfeit protection, identification, and/or data embedding using DNA sequences.

BACKGROUND OF THE INVENTION

[0003] There is a demand for reliable, durable, and accurate methods of identification and authentication in a multitude of markets. These markets could include, but are not limited to, luxury items, collectibles, artworks, wine, spirits, raw materials (such as raw minerals, processed minerals, intermediate materials), currency, or any other physical object where inherent embedding of identification, authenticity, data, and/or traceability information is desired. Counterfeit goods lead to loss of revenue, damage to reputation, brand dilution, and circumvent safety and sustainability standards. Proper authentication methods allow for tracing of importation/exportation of goods and/or verification of object provenance. Previous approaches to address this demand include incorporation of extrinsic markers in product packaging or on the product itself. Extrinsic markers include watermarks, holograms, serialization marks, engravings, microprinting, smart labels (e.g., QR codes), specialty inks, guilloche patterns, and microscopic coatings (e.g., dust identification). However, extrinsic markers are still amenable to counterfeit. Alternatively, intrinsic markers, embedded in the product, have been developed to further increase the difficulty of counterfeiting efforts, such as radio frequency identification (RFID) tags, near field communication (NFC) tags, spectral and/or isotopic fingerprints, and blockchain tracking. Unfortunately, intrinsic markers are limited in their application and may become more vulnerable as technology develops.

[0004] It is known that DNA can be encapsulated in nanometer silica beads, which can be fused into various materials that are used to print or cast objects in any shape and subsequently recovered.

See, e.g., Koch J, et al., “A DNA-of-things storage architecture to create materials with embedded memory.” Nat. Biotechnol. (2020)38(1):39-43; e.g., U.S. Patent No. 9,850,531, “Molecular code systems”; e.g., Bossert, et al., “A hydrofluoric acid-free method to dissolve and quantify silica nanoparticles in aqueous and solid matrices” Sci. Rep. (2019)9:7938, the contents of each of which are incorporated herein by reference. However, simply embedding or incorporating identifying DNA into an object is not an effective approach to counterfeit prevention, if the DNA can be readily retrieved, amplified, and embedded into counterfeit goods.

[0005] Furthermore, the synthesis and retrieval of data stored in DNA can be time, resource, and financially costly. Methods developed to optimize time, reagent use, and decoding efficiency may provide improved methods of DNA synthesis and counterfeit protection. Moreover, DNA data storage using single-base accuracy allows for high-density data storage, but also requires additional time and material resources to both encode and retrieve user-defined data. Such processes may limit the quality and quantity of synthesized DNA. However, methods have been developed wherein data encoding does not require single-base accuracy. See, e.g., Lee, H.H., *et al.*, “Terminator-free template-independent enzymatic DNA synthesis for digital information storage.” Nat. Commun. (2019)10:2383, the contents of which are incorporated herein by reference.

[0006] There remains a need for improvement regarding methods of authentication and counterfeit protection of goods.

BRIEF DESCRIPTION OF THE INVENTION

[0007] DNA can prove a useful material for object authentication and object provenance, wherein data is encoded within one or more DNA sequence, incorporated into an object of interest, and is subsequently removed and analyzed. Analysis of such DNA sequences may provide a “fingerprint;” for example, various methods such as (A+T)/(G+C) ratio determination, restriction fragment length polymorphism (RFLP), mass spectrometry (MS), and DNA sequencing produces production fingerprints that allow for the detection, tracking, and/or authentication of one or more DNA sequences incorporated into an object. Indeed, the benefits and applications of the effectively infinite design space afforded by combinations of DNA markers has been discussed in our previous work, e.g., U.S. Patent Application Nos. 63/582,199 and 63/623,085, the contents of which are incorporated herein by reference.

[0008] This disclosure is directed, in one aspect, to a novel population of deoxyribonucleic acid (DNA) sequences encoding data useful in the authentication of objects and for protection against counterfeiting, comprising nucleic acid data packets (“nackets”), wherein each nacket contains a plurality of DNA molecules encoding the same data, wherein the sequences of the DNA molecules are heterogeneous. The nackets may also be referred to herein as DNA (or polymer) memory strings or memory strands. For example, the nackets may be prepared by heterologous (or heterogeneous or varied) cassette data writing, wherein two or more cassette sequences are provided for (or associated with or indicative of) a single bit or combination of bits in a machine-readable code, e.g., a binary code, such that all or nearly all the DNA molecules in the nacket encode the same data, but the sequences of the individual molecules exhibit extremely high variation, e.g., due to the use of heterologous cassettes encoding the same bit or bits of data, e.g., wherein the percent abundance of the different cassette variants used in writing the nackets provides a unique and distinguishable feature of the nacket.

[0009] In some embodiments, the nackets are synthesized using one or more transferase enzyme, e.g., terminal deoxynucleotidyl transferase (TdT). For example, the nackets may be prepared by stepwise addition of non-identical nucleotides forming homopolymer extensions within the DNA sequence, wherein the transition from a first homopolymer extension to a second homopolymer extension comprises a transition between non-identical nucleotides, and wherein the transition(s) between non-identical nucleotides provide for (or are associated with or indicative of) a single bit or combination of bits in a machine-readable code, e.g., a ternary code, such that a population of DNA molecules encodes a desired data string.

[0010] In some embodiments, the nackets are synthesized using topoisomerase mediated ligation. For example, synonymous cassettes, having different sequences but encoding the same information, can be added in each addition step, to build a set of DNA polymers, wherein each polymer has a series of informational cassettes encoding substantially the same information but wherein the polymers are heterogeneous at a sequence level.

[0011] The nackets may be incorporated into or associated with goods for purposes of identifying and authenticating the goods. In certain embodiments, the nackets are adsorbed to (or encapsulated within) silica beads or particles, which are optionally coated with polymer, and incorporated into goods, e.g., for purposes of identification and authentication of the goods. In certain embodiments, the nackets are added to an ink, e.g., a water-soluble ink, optionally comprising a polymer, e.g.,

for purposes of identification and authentication of signatures, documents, and prints. In certain embodiments, the nackets are adsorbed to (or encapsulated within) silica beads or particles after synthesis or production of the nackets. In alternative or additional embodiments, the nackets are adsorbed to (or encapsulated within) silica beads or particles during synthesis or production of the nackets, e.g., during a one-pot synthesis of the nackets and silica beads or particles. In certain embodiments, the nackets are integrated into ceramic or silica beads or particles using a sol-gel process comprising reacting a molecular precursor (e.g., a silicate, for example tetraethylorthosilicate) with water in an alcoholic solution comprising the nackets, and condensing the product to form a crosslinked particle structure containing the nackets within the cross-linked structure, e.g., a Stöber nanoparticle reaction.

[0012] In another aspect, the disclosure is directed to methods of marking, identifying, and authenticating goods, comprising (i) marking the goods by incorporating or associating the nackets described herein with the goods to be identified or authenticated, and (ii) identifying and authenticating the goods thus marked, by retrieving and sequencing the nackets, identifying the goods based on the data, e.g., binary coded data, e.g., ternary coded data, encrypted in the nackets, and authenticating the goods by (i) measuring the relative amounts of the different cassette variants and/or (ii) analyzing the DNA sequence(s), e.g., a DNA “fingerprint”, e.g., transitions between non-identical nucleotides, (iii) and/or sequencing and decoding the coded data.

[0013]

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] Figure 1 schematically depicts a process for topoisomerase mediated ligation using DNA cassettes with complimentary overhangs, and 5' phosphate and phosphatase for blocking and de-blocking, to permit controlled, single cassette additions.

[0015] Figure 2 depicts a DNA molecule comprising cassettes ligated by the process depicted in Figure 1.

[0016] Figure 3 illustrates the potential for a high degree of diversity in topogation cassettes.

[0017] Figure 4 depicts two-bit, multi-base encoding as opposed to two-bit, single base encoding.

[0018] Figure 5 illustrates how a very high diversity of combinations can be generated using multi-base encoding (heterologous cassettes) for a production fingerprint (i.e., signature of the manufacturing process).

[0019] Figure 6 shows examples of cassettes useful for homologous cassette data writing and for heterologous cassette data writing, using two unique, non-interacting overhangs (A and B), to permit addition of one cassette in each reaction.

[0020] Figures 7-13 show schematically how the heterologous cassette data writing generates a unique mixture of DNA.

[0021] Figure 14 shows advantages of heterologous cassette data writing compared to single base writing.

[0022] Figure 15 provides an example of how a 32-byte NFT could be encoded into 16, 12-cassette chains.

[0023] Figure 16 provides an overview of preparing the nackets and incorporating them into products.

[0024] Figure 17 provides an overview of retrieving and analyzing the nackets to verify authenticity.

[0025] Figure 18 provides a schematic overview of different roles in the verification process.

[0026] Figure 19 is a diagram showing topo cassettes representing various combinations of binary bits, in accordance with embodiments of the present disclosure.

[0027] Figure 20 is a diagram showing the number of potential topo cassettes based on the number of positions and number of different DNA bases, in accordance with embodiments of the present disclosure.

[0028] Figure 21 is a diagram showing how multiple different cassettes may be used to specify the same underlying binary information, in accordance with embodiments of the present disclosure.

[0029] Figure 22 is a diagram showing a comparison of homogeneous cassette data writing and heterogeneous cassette data writing using a plurality of topo cassettes combined in a predetermined formulation or mixture, in accordance with embodiments of the present disclosure.

[0030] Figure 23 is a diagram showing the heterogeneous mixtures of topo cassettes of Fig. 22 loaded into print heads of a ink jet DNA printer, in accordance with embodiments of the present disclosure.

[0031] Figure 24 is a diagram showing a process for writing two-bit binary codes onto the surface of a substrate or matrix, in accordance with embodiments of the present disclosure.

[0032] Figures 25A, 25B, 25C, 25D, 25E, 25F, 25G, 25H, 25I, and 25J are diagrams showing a process for writing memory strings at a spot on a substrate using a pre-set formulation or mixture of cassettes for each 2-bit pair, in accordance with embodiments of the present disclosure.

[0033] Figure 26 is a diagram showing a cassette along a memory string and cassettes assigned to each 2-bit code in the memory string, in accordance with embodiments of the present disclosure.

[0034] Figures 27A, 27B, 27C, and 27D are diagrams showing a process for validating memory strings or nackets using the predetermined cassette mixture associated with a given 2-bit binary code, in accordance with embodiments of the present disclosure.

[0035] Figure 28 is a diagram showing two dimensions of randomness and validation of memory strings (or nackets) along a memory string and across all memory strings for a given spot, in accordance with embodiments of the present disclosure.

[0036] Figures 29A, 29B, and 29C are tables showing various assignments between binary codes and cassettes and associated cassette mixtures/formulations, based on lot numbers, in accordance with embodiments of the present disclosure.

[0037] Figure 30A is a block diagram showing an inkjet printing system showing print head control and wafer array/stage control logic and an instrument for fluidics/reagents, in accordance with embodiments of the present disclosure.

[0038] Figure 30B is a block diagram of a computer system of Figure 30A, in accordance with embodiments of the present disclosure.

[0039] Figure 31A is a flow diagram for writing (printing) and unloading coded polymer memory strings in an inkjet writing system, in accordance with embodiments of the present disclosure.

[0040] Figure 31B is a flow diagram for writing (printing) 2-bit code to DNA/polymer memory string in an inkjet writing system, in accordance with embodiments of the present disclosure.

[0041] Figure 32A is a side view diagram showing several spots with coded DNA and cleaving fluid for removing coded DNA strands from surface of substrate, in accordance with embodiments of the present disclosure.

[0042] Figure 32B is a diagram showing an array of spots with coded DNA having columns (X) of redundant spots with the same encoded DNA data written, and rows (Y) of spots with different encoded DNA written, in accordance with embodiments of the present disclosure.

[0043] Figure 33 is a diagram showing removal of spotted DNA from surface of substrate to a collection bin and reading and decoding the DNA collection, in accordance with embodiments of the present disclosure.

[0044] Figure 34 is a flow diagram for decoding and confirming polymer memory string data, in accordance with embodiments of the present disclosure.

[0045] Figures 35A and 35B are diagrams showing examples of cassettes making up address, data, and error checking for written DNA/polymer memory strings, in accordance with embodiments of the present disclosure.

[0046] Figure 36A is a diagram showing a method for creating unique cryptographic DNA fingerprints, in accordance with embodiments of the present disclosure.

[0047] Figure 36B is a diagram showing three layers of data derived from a common DNA sequence, in accordance with embodiments of the present disclosure.

[0048] Figure 37 is a diagram showing a method for encoding/decoding system for encoding and decoding a digital file to and from DNA, in accordance with embodiments of the present disclosure.

[0049] Figures 38A, 38B, 38C, 38D, 38E, 38F are diagrams showing a method for the system of Fig. 37 for encoding a digital file into DNA for writing, in accordance with embodiments of the present disclosure.

[0050] Figures 39A, 39B, 39C, 39D, 39E, 39F, 39G are diagrams showing a method for the system of Fig. 37 for decoding written DNA back into the original digital file, in accordance with embodiments of the present disclosure.

[0051] Figures 40A, 40B, 40C are data graphs showing results data using the encode/decode system of Fig. 37, in accordance with embodiments of the present disclosure.

[0052] Figure 41 schematically depicts a trit encoding map or schema.

[0053] Figure 42 schematically depicts the variable space of unique DNA sequences synthesized using heterologous DNA cassette data writing.

[0054] Figure 43 displays DNA stability and recovery at 2 and 6 weeks after being written on paper using fountain pen ink.

[0055] Figure 44 displays DNA stability and recovery at 8 weeks after being written on paper using fountain pen ink.

[0056] Figure 45 displays the variable space of unique DNA sequences while being synonymous in the encoding of a NFT code.

[0057] Figure 46 displays recovery efficiency of DNA after accelerated aging of samples written on paper using fountain pen ink.

[0058] Figure 47 displays the relative frequency of double-strand DNA breakage during accelerated aging of samples written on paper using fountain pen ink.

[0059] Figure 48 displays a relatively stable sequence error rate throughout accelerated aging of samples written on paper using fountain pen ink, while sequence efficiency decreases over time.

[0060] Figure 49 displays the shift of sequence length distribution over time.

[0061] Figure 50 is a diagram showing print head banks for a laser jet DNA printer having separate topo cassettes nozzles within a head bank, and having multiple head banks, in accordance with embodiments of the present disclosure.

[0062] Figure 51A is a diagram showing an array of spots with coded DNA on a chip/array having rows (Y) of spots with different encoded DNA written, each row having computer-generated random proportions of cassettes (Cs) associated with each two-bit code, in accordance with embodiments of the present disclosure.

[0063] Figure 51B is a diagram showing an array of spots with coded DNA on a chip/array, the entire chip having computer-generated random proportions of cassettes (Cs) associated with each two-bit code for a given lot number, in accordance with embodiments of the present disclosure.

[0064] Figure 52 is a diagram showing print head banks for a laser jet DNA printer having separate topo cassettes nozzles within a head bank, in accordance with embodiments of the present disclosure.

[0065] Figure 53A is a flow diagram for writing (printing) and unloading coded polymer memory strings in an inkjet writing system using computer-based randomness for cassette writing selection, in accordance with embodiments of the present disclosure.

[0066] Figure 53B is a flow diagram for writing (printing) 2-bit code to DNA/polymer memory string in an inkjet writing system using computer-based randomness for cassette writing selection, in accordance with embodiments of the present disclosure.

[0067] Figure 54 is a flow diagram for decoding and confirming polymer memory string data when using computer-based randomness for cassette writing selection, in accordance with embodiments of the present disclosure.

DETAILED DESCRIPTION OF THE INVENTION

[0068] The following description of different embodiments is merely exemplary in nature and is in no way intended to limit the invention, its application, or uses.

[0069] We have previously described information storage using a charged polymer, for example DNA, comprising at least two distinct monomers or oligomers, wherein information is encoded in a machine-readable code, for example a binary code. For example, US 11505825, US 11655465, and U.S. Application No. 18/358,861, filed July 25, 2023, each incorporated herein by reference, describe, among other things, methods of synthesizing a DNA molecule using topoisomerase-mediated ligation, adding informational cassettes to a DNA strand in the 3' to 5' direction.

[0070] The following commonly-owned issued patents contain subject matter related to that described herein, each of which are hereby incorporated by reference in their entirety to the fullest extent permitted by applicable law: US Patent 10,438,662 and US Patent 10,640,822. The aforementioned commonly-owned patents discuss approaches for writing (or storing) data in a charged polymer, e.g., DNA, using Add "0" and Add "1" enzymes and a deblock enzyme, as described therein.

[0071] The following commonly-owned US Patent Application Ser. Nos. 18/358,861 and 18/444,662 contain subject matter related to that described herein, which are hereby incorporated by reference in their entirety to the fullest extent permitted by applicable law. The aforementioned commonly-owned patent applications discuss other approaches for writing (or storing) data in a charged polymer, e.g., DNA, such as, using an AB Adapter instead of a deblock enzyme and using "A0B" and "A1B" for the Add "0" and Add "1" reagents, as described therein, and methods of writing strands of DNA cassettes using inkjet reaction formats.

[0072] As discussed herein, the present disclosure provides a novel system of storing (or writing or printing) information (or data) using a charged polymer, e.g., DNA, the monomers of which correspond to a machine-readable code, e.g., a binary, ternary, or other base code, and which can be synthesized in various ways, including using a piezo-electric inkjet printer system, such as that discussed in US patent application No. 18/444,662, filed Feb. 17, 2024, which is incorporated herein by reference in its entirety to the fullest extent permitted by applicable law.

[0073] Topoisomerases are enzymes that spontaneously recognize and cleave at least one strand of a double strand of nucleic acids within a sequence segment known as the site-specific recombination sequence. Vaccinia topoisomerase is a type I DNA topoisomerase that has the ability to cut DNA strands 3' of its recognition sequence of 5'-(C/T)CCTT-3', e.g., 5' CCCTT 3', and to ligate, or rejoin the DNA back together again. Oligonucleotide cassettes containing digital information can be linked together by topoisomerases. In this approach, the DNA base cassette contains a topoisomerase recognition sequence, thereby allowing it to be "charged" with a topoisomerase, such that a strand of DNA is cleaved by the enzyme, and becomes transiently covalently bound to a topoisomerase at the 3' end. When an appropriate DNA acceptor is found, the topoisomerase ligates the cassette to the DNA acceptor strand in a process referred to as "bit addition" or "topogation". After ligating the DNA cassette onto a DNA acceptor strand, the topoisomerase is no longer bound to the DNA.

[0074] Figure 1 depicts a process for topoisomerase mediated ligation using DNA cassettes with complimentary overhangs, and 5' phosphate and phosphatase for blocking and de-blocking, to permit controlled, single cassette additions. In alternative embodiments, blocking and de-blocking may be accomplished using thermally-reactive moieties, light-reactive moieties, enzyme-reactive moieties, or combinations thereof. In a simple embodiment, there are two pools of cassettes, which can be added one by one to the DNA strand to provide a binary code sequence, e.g., X or Y. So if sequence X = 1 and sequence Y = 0, a binary sequence 1001 can be encoded by forming a strand comprising a series of cassettes X – Y – Y – X. Each cassette may further comprise a spacer region, and/or the cassettes may be separated by one or more spacer regions, wherein the spacer regions may comprise a topoisomerase recognition sequence and a short complementary sequence, as relics of the topogation process, as depicted in Figure 2. The cassettes can contain multiple bits (e.g., XX, XY, YX, YY) to allow building an informational sequence with fewer operations. But in these cases, the pools from which the cassettes are taken are homogeneous – all the "X"s have a characteristic sequence, and all the "Y"s have a different characteristic sequence, for example.

[0075] In the present disclosure, multiple defined sequences encode a specific bit or combination of bits. Topoisomerase cassettes can be highly variable. As depicted in Figure 3, cassettes of varying length and base composition can be made to encode the same or different bits. While the linker sequences are conserved, the sequences used to convey information need not be. For example, bit X may be encoded by different sequences X1, X2, X3, or X4, and bit Y may be

encoded by Y1, Y2, Y3, or Y4. This permits heterogenous cassette data writing, so that a very large number of different sequences can encode the same data. This permits multiple layers of information – the binary code information lies on top of a more complex mixture of sequences, allowing layered data that lends itself to product identification. For example, in identifying a product, the first layer of data could be considered to be the product's appearance and label (fairly easy to replicate), the second as the binary code encoded by the series of cassettes (somewhat more difficult to replicate), and the third as the precise mixture of the heterologous cassettes used to encode the binary data (far more difficult to replicate). Examples are depicted in Figures 4 and 5.

[0076] In particular, Figure 5 shows two different cassette formulations or mixtures (mix1, mix2). Referring to Figures 4 and 5, when 2-bit binary encoding is used, each two-bit combination can be represented by Y different cassettes simultaneously in specific formulations. Sequences can be formulated in varying ratios for additional combinatorial complexity. For example, $(100^Y)^4$ formulations are possible, assuming integer percentages of each potential sequence in a formulation. Also, with N coding bases (or positions) in each 15-base representation, 4^N or 4^{15} variants are possible if all 15 base positions are used.

[0077] Figure 6 shows examples of cassettes useful for homologous cassette data writing and for heterologous cassette data writing, using two unique, non-interacting overhangs (A and B), such that A overhangs (CACT on the top strand and GTGA on the bottom) are complementary, and B overhangs (GGCA on the top and CCGT on the bottom) are complementary, but the A overhangs are not complementary with B overhangs, thereby permitting addition of one cassette in each reaction, without the need for a protection/deprotection, as generally described in U.S. Application No. 18/358,861. In this system, four cassettes are needed to provide a binary (0,1) code, e.g., A0B, A1B, B0A, and B1A. But in the heterologous cassette data writing example, there are two different informational sequences for 1 and two for 0, so there are a total of eight different cassettes. Moreover, the proportion of these cassette types can be varied (e.g. 50%/50% or 25%/75% as depicted), resulting in DNA sequences that have the same binary code information, but different sequences and DNA populations having different proportions of the different cassettes.

[0078] Using heterologous cassette data writing permits significant opportunity for identification and counterfeit protection. Each data writing fluid contains two or more unique cassette sequences that are distinct across the writing set, e.g., D, d, M, m, as in Figure 6. The data represented by the cassettes in a given fluid can be same, e.g. for copy protection / counterfeit protection and to enable

reading on short read sequencers, or different, for creating a cassette based UMI or any random number generation (e.g. random number applications). The sequences of cassettes can be shortened to a single letter, where the case of the letter represents AB (lower case) or BA (upper case). In Figure 6, this is demonstrated by D, d, E, and e all representing 0 and M, m, N, and n all representing 1. Thus, you can shorten a complex DNA sequence dramatically and ease visual interpretation of the results. Further, the standard handling of sequence files (e.g., FASTA, FASTQ, string manipulations, matching, etc.) are then all compatible with this “Data Sequence” notation enabling a vast, mature toolset amenable to the heterogeneous data layer. For the purpose of data writing, one convention, for example, is that the first letter in a set of letters is used in the encoding phase to represent which fluid is used to write the associated nacket. All symbols are used during decoding in a process where software finds the best match between a component sequence and the most relevant “data sequence” letter. The incidence rate of each can be controlled by writing, based on the relative amounts of the different cassettes, and then measured from sequencing. This incidence rate can be used as a unique fingerprint of the reagents used to write the data. One could also encode data into the levels of each in the fluid ratios (e.g. a lot code / etc.). A fingerprint could be obtained in addition to the lot coding.

[0079] In certain embodiments, one or more cassettes are synthesized using sequential single-base addition methods, e.g., phosphoramidite synthesis. In certain embodiments, one or more cassettes are synthesized using enzymatic methods, e.g., one or more DNA polymerase, e.g., one or more flap endonuclease, e.g., one or more DNA ligase, e.g., one or more topoisomerase. In certain embodiments, one or more cassettes are synthesized using sequential single-base addition and/or enzymatic methods before amplification of the cassettes to provide a larger yield of DNA production, e.g., amplification using PCR (polymerase chain reaction), e.g., amplification using RCA (rolling circle amplification). In certain embodiments, one or more cassettes are ligated together using methods comprising single-base addition techniques, e.g., phosphoramidite chemistry, enzymatic methods, e.g., DNA polymerase, e.g., flap endonuclease, e.g., DNA ligase, e.g., topoisomerase, or a combination thereof. In certain embodiments, the one or more cassettes are synthesized using non-natural nucleotides or nucleobases. In certain embodiments, the one or more cassettes are further modified after synthesis, optionally after ligation to one or more other cassettes, e.g., modified with small molecule moieties, polymers, click-active reagents, fluorescent markers, etc.

[0080] Figures 7-13 show schematically how the heterologous cassette data writing generates a unique mixture of DNA. In this example, the binary data for all molecules of the nucleic acid data packet (“nacket”) is 011011, where each cassette represents a single binary bit (0,1). But due to cassette heterogeneity in the writing fluids, all molecules written are unique in the same nacket: !EnMeNn#, !EnNdMm#, !EnNeNn#, !EnNeNm#, !DmNeMn#, !DmNeNn#, !DmNdNm#, and !DnMeNn# are the sequences for the eight molecules generated, where “!” is a starter string or acceptor string and “#” is an end cap at the end of the nacket or memory string. The starter string and ending string may include other features useful for data storage or authentication; for example, unique “primer regions” may be included in these zones. The number of permutations is approximately: (# of unique chains or cassettes per fluid) ^ (# of rounds of cassette addition). For this example, with 6 rounds of addition (i.e., 6 cassettes) with 2 unique chains (or cassettes) per fluid, there are 2^6 or 64 unique molecules permutations for each nacket. For a 150-cassette chain with 4 unique chains (or cassettes) per fluid: $\sim 4^{150}$, or about $2e90$ unique molecules for each nacket.

[0081] Each read of this nacket generates three layers of data:

- a. Nacket Data Layer (here, 011011): One value per nacket ID.
- b. Production Lot Fingerprint: Measurements of the percent abundance of the different cassette variants used in writing. The original fingerprint can be stored on a block chain.
- c. Object Fingerprint: A list of random sequences from each read. Here the decoding sequence has unique values. A certain number of numbers during verification reading must match those originally found.

This presents a number of advantages:

- All three layers of data are in the same DNA sequence – they are inseparable.
- The top layer enables the sequence to contain digital data, which may be tied to a block chain, one or more elements of a public-key infrastructure, a digital identifier to any proprietary or public information system, and/or any amount of digital data.
- The external systems, such as a block chain, public-key infrastructure system, or other data system may contain information to validate the other two.
- Permits use of public block chain, which will survive even if the company synthesizing the DNA goes out of business.

- There are many techniques to sequence DNA. Other systems may come and go, but DNA will always be readable.

[0082] Creating the DNA using ligation, e.g., topogation, of a series of cassettes rather than single base addition creates significant advantages because of the longer chain length and permutation space. Figure 14 shows the impact on synthesis yield varying single base chemical coupling efficiencies compared to cassette data writing as described herein. Using single base synthesis, blocks of less than 6 base pairs are highly susceptible to sequencing errors, ligation yield and are susceptible to counterfeit, whereas blocks of greater than seven base pairs fall outside desirable yields for all single base synthesis chemistries. Long sequences or sets of sequences of DNA could be prepared, e.g., using amplification in PCR or phages and used as an identifying marker, but such a marker would lend itself to counterfeiting, because the sequence or sequences could be readily isolated, amplified, and applied to fake goods.

[0083] The nackets described herein are particularly suitable for efficient analysis by conventional DNA sequencers, such as short-read sequencers and/or long-read sequencers, such as Illumina sequencers. One of skill in the art will readily appreciate the benefits of each approach, and the situations wherein short-read sequencing and/or long-read sequencing is most appropriate; e.g., short-read sequencing for nackets comprising 12 or less cassettes, and long-read sequencing for nackets comprising greater than 12 cassettes. The nackets are about two to six kilobases long and have repeating sequences across many data chains due to the reuse of cassettes. Each cassette is about 20 bases long, meaning about 100-300 cassettes fit in one typical read. Using heterogeneous cassette data writing (e.g. 4 flavors of cassette per data writing fluid), every chain would be fully unique prior to amplification. After amplification, some number would be “selected” and enriched.

[0084] In certain embodiments, an encoding scheme compatible with short-read sequencers is used. For example, nackets comprising 10 to 12 cassettes, wherein each cassette comprises about 20 base pairs in length, and about 30 base pairs for each of the integral short read sequencing primers in the starting and ending strands, yields nackets of 260 to 300 base pairs in length. Such nackets would be readily compatible with a variety of short-read sequencers. In this scenario, in order to obtain sufficient complexity of the fingerprint, the heterogeneity must be larger than 4 for the typical application. For example, with a heterogeneity of 10, this yields 10^{10} to 10^{12} unique permutations. Thus, applications that use long-read sequencers may provide an advantage in reading molecules with more variations.

[0085] In certain embodiments, the nackets may be analyzed using “rapid fingerprinting” techniques. In certain embodiments, rapid fingerprinting provides for initial evaluation of nackets that does not require sequencing of the full nacket sequences. In certain embodiments, rapid fingerprinting yields analytical results in less than 30 minutes, e.g., in less than 15 minutes, e.g., in less than 10 minutes, e.g., in less than 5 minutes, e.g., in less than 3 minutes, e.g., in less than 2 minutes, e.g., in less than 60 seconds, e.g., in less than 50 seconds, e.g., in less than 45 seconds, e.g., in less than 40 seconds, e.g., in less than 35 seconds, e.g., in less than 30 seconds, e.g., in less than 25 seconds, e.g., in less than 20 seconds, e.g., in less than 15 seconds, e.g., in less than 12 seconds, e.g., in less than 10 seconds, e.g., in less than 9 seconds, e.g., in less than 8 seconds, e.g., in less than 7 seconds, e.g., in less than 6 seconds, e.g., in less than 5 seconds. In certain embodiments, rapid fingerprinting comprises exposing the nackets to fluorescent probes, azide-alkyne cycloaddition reagents, antibodies, microsatellites, or a combination thereof, and/or through use of restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), or a combination thereof. In certain embodiments, a chip platform, e.g., a nanochannel or microchannel array, is provided comprising the complementary reagents necessary to perform rapid fingerprinting, for example, for field-deployable analysis. In certain embodiments, the chip platform comprises capture sequences and/or PCR primer sequences that are complimentary to the nackets and allow for subsequent identification, optionally comprising amplification of said nackets. In certain embodiments, the nackets comprise terminal nucleotide sequences or DNA “caps” which allow for capture/sequestration, binding of the nackets to the chip platform, and subsequent identification.

[0086] In one embodiment, to enable rapid fingerprinting, during writing, a mixture of starter molecules and/or ending molecules may be used, wherein each has a unique primer sequence that is identifiable via rapid nucleic acid amplification test (NAAT). This may be a single target to prove presence, or a complex fingerprint of molecules. Each set of starter molecules and/or ending molecules may be mixed and associated with the authentication data either directly or through a hashing function. In one embodiment, this may comprise 32 unique starter molecules that all attach to the surface and accept the first topogation reaction, but will react with different primers in a NAAT test. When a sample is obtained and reacted with this NAAT test, a fingerprint of 32 YES/NO answers may be produced, which yields a 32 bit unique ID or 4 billion unique combinations. That ID would be different for every writing process. In another embodiment, this

could be done with 32 starter molecules and 32 ending molecules, yielding 64 bits of 1.8×10^{19} permutations or possibilities.

[0087] The nackets may encode a non-fungible token (NFT), which is a unique digital identifier that is recorded on a blockchain, and is used to certify ownership and authenticity. It cannot be copied, substituted, or subdivided. Figure 15 provides an example of how a 32-byte NFT could be encoded into 16, 12-cassette chains. Figure 16 provides an overview of preparing the nackets and incorporating them into products. Figure 17 provides an overview of retrieving and analyzing the nackets to verify authenticity. Figure 18 provides a schematic overview of different roles in the verification process.

[0088] In particular, referring to Figure 16, in some embodiments, a first step is to mint the NFT, or create blockchain NFT token and binary code, which may use the public blockchain or private blockchain. Next, step 2, is to synthesize the DNA chains or strings with the binary encoding as discussed herein, which may include blockchain NFT token, production metadata and cryptographic fingerprinting. Next, step 3 may be encapsulation of the DNA into a material, such as silica beads or plasmids. In particular, DNA is in a stable dried form, silica further stabilizes DNA, optical properties of objects are unaffected by beads, they are safe for human consumption, and beads can be extracted from materials and the DNA sequenced, and plasmids can be put into living organisms if desired. Also, plasmids with DNA codes can be easily transfected into bacteria, cells, plants, animals, or fungi. Next, step 4 is to embed the beads or the like into the desired objects. Next, referring to Figure 17, step 5 is to sample the object with the embedded beads (or the like). Next, step 6 is to extract the beads with DNA from the object and elute the DNA chains or strings. In particular, this step is to extract silica beads or plasmids and isolate DNA chains using known and robust processes for bead extraction from materials, and elution of DNA from beads is well known, characterized and published, and plasmid extraction is also well known by those skilled in the art. Next, step 7 is to sequence the extracted DNA chains. The DNA may be read with any known commercial sequencer (e.g., made by Illumina, oxford nanopore, or others), and the cassettes may be designed for peak performance in any sequencing chemistry, and can also leverage a global network of commercial sequencing labs for third party sequencing. Next, step 8 is to verify the DNA encoded binary codes, which may be in the form of an NFT or NFT hash. In particular, this step may verify the presence of the blockchain NFT token, production metadata, and/or cryptographic fingerprint, as applicable.

[0089] The nackets may encode one or more public-private key infrastructure elements, which may be pulled from private and/or public certificate authorities. The use of the certificate authority can be used to mediate the validity of the underlying object, for example, by revoking the associate certificate if the object is known to be stolen. Authenticity information may be further stored in a public information system, wherein said information may be accessed online, for example, using a PKI infrastructure to validate the authenticity of the remote server being used to validate the physical object.

[0090] This disclosure is directed, in another aspect, to a nucleotide polymer, e.g., deoxyribonucleic acid (DNA), synthesized in a *de novo* enzymatic process using terminal deoxynucleotidyl transferase (TdT). TdT is a template-independent polymerase that extends an “initiator” strand of DNA by the addition of one or more deoxyribonucleotide triphosphate (dNTP) monomers onto the 3’ terminus of said initiator strand. Apyrase is an enzyme that mediates nucleic acid substrate degradation, wherein apyrase degrades nucleoside triphosphates into the corresponding diphosphate or monophosphate precursors; said precursors are TdT-inactive. By optimizing the relative concentrations of TdT and apyrase within a reaction mixture, these enzymes can be made to compete against one another such that stepwise addition of dNTPs onto the 3’ terminus of DNA initiator strands can be kinetically-controlled. Thus, through iterative addition of dNTPs onto one or more initiator strands, DNA strands with short homopolymeric extensions are produced wherein data, e.g., user-defined data, are encoded within a nucleotide polymer, e.g., DNA, producing nucleic acid data packets (“nackets”). Using this approach, data are not encoded in the specific nucleotide sequence *per se*, but rather the data are encoded in the transitions between non-identical nucleotides within the polymer.

[0091] In some embodiments, initiator strands are placed in contact with a reaction mixture comprising TdT and apyrase, wherein dNTP monomers are introduced to said reaction mixture in iterative, stepwise additions of non-identical dNTP species. In some embodiments, dNTP species comprise adenosine triphosphate (ATP), guanosine triphosphate (GTP), cytidine triphosphate (CTP), thymidine triphosphate (TTP), and optionally uridine triphosphate (UTP). In some embodiments, ATP, GTP, CTP, TTP, and UTP may be referred to by their corresponding nucleobases, i.e., A, G, C, T, and U, respectively; those of skill in the art will readily understand the use of nucleobase terms to describe nucleotides in various available phosphorylated states based on the context in which the nucleobase terms are used. For example, if a first addition step

consists of the addition of A, i.e., adenosine triphosphate, onto the 3' terminus of a DNA strand within a reaction mixture, the next stepwise addition may comprise, e.g., G, C, or T, but said next stepwise addition may not be the addition of A since this would not encode any additional information onto the DNA strand relative to the first addition step.

[0092] In some embodiments, the reaction mixture, DNA initiator strands, and dNTPs come into contact under flow conditions. In some embodiments, the reaction mixture, DNA initiator strands, and dNTPs come into contact under mixing conditions. In some embodiments, the reaction mixture, DNA initiator strands, and dNTPs come into contact in solution, e.g., droplet or bulk solution, e.g., without active mixing.

[0093] In some embodiments, the stepwise addition of dNTPs onto the 3' terminus of a DNA initiator will produce homopolymer extensions of heterogeneous lengths. For example, within a single addition reaction step, a first DNA strand may be extended by one or more dNTP monomers, e.g., 2 dNTP monomers, while a second DNA strand may be extended by one or more dNTP monomers, e.g., 3 dNTP monomers. In some embodiments, homopolymer extensions of a DNA strand may comprise 1 or more dNTP additions, e.g., 2 or more dNTP additions, 3 or more dNTP additions, 4 or more dNTP additions, 5 or more dNTP additions, 6 or more dNTP additions, 7 or more dNTP additions, 8 or more dNTP additions, 9 or more dNTP additions, 10 or more dNTP additions, 15 or more dNTP additions, 20 or more dNTP additions, 25 or more dNTP additions, 30 or more dNTP additions, 35 or more dNTP additions, 40 or more dNTP additions, 45 or more dNTP additions, 50 or more dNTP additions, etc. In some embodiments, homopolymer extensions of a first DNA strand are independent from homopolymer extensions of a second, third, fourth, etc., DNA strand.

[0094] In some embodiments, the synthesis reaction produces a population of synthesized strands comprising a series of homopolymer extensions of heterogeneous lengths. In some embodiments, the population of synthesized strands all comprise the same number and sequence of nucleotide transitions between the homopolymer extensions, while said homopolymer extensions are of heterogeneous lengths.

[0095] In some embodiments, the reaction mixture may comprise aqueous conditions. In some embodiments, the reaction mixture may comprise buffer conditions. In some embodiments, the reaction mixture may comprise further additives, e.g., ions, e.g., cations, e.g., divalent cations, e.g., cobalt.

[0096] In some embodiments, the reaction mixture may comprise a ratio of TdT to apyrase of about 10,000:1 to about 100:1, e.g., the reaction mixture may comprise a ratio of TdT to apyrase of about 5,000:1 to about 500:1, e.g., 4,000:1 to about 800:1, e.g., 4,000:1 to about 1,000:1; e.g., about 4,000:1, or about 1,000:1. In some embodiments, the reaction mixture may comprise a concentration of TdT of about 0.1 U/ μ L to about 10 U/ μ L, e.g., about 0.5 U/ μ L to about 5 U/ μ L, e.g., about 0.7 U/ μ L to about 3 U/ μ L, e.g., about 0.8 U/ μ L to about 2 U/ μ L, e.g., about 0.9 U/ μ L to about 1.5 U/ μ L, e.g., about 1 U/ μ L to about 1.2 U/ μ L, e.g., about 1 U/ μ L. In some embodiments, the reaction mixture may comprise a concentration of apyrase of about 0.1 mU/ μ L to about 10 mU/ μ L, e.g., about 0.1 mU/ μ L to about 5 mU/ μ L, e.g., about 0.2 mU/ μ L to about 2 mU/ μ L, e.g., about 0.25 mU/ μ L to about 1.5 mU/ μ L, e.g., about 0.25 mU/ μ L to about 1 mU/ μ L, e.g., about 0.25 mU/ μ L, or about 1 mU/ μ L.

[0097] In some embodiments, the reaction mixture comprises dNTPs, e.g., dATP, dCTP, dGTP, and/or dTTP. In some embodiments, the reaction mixture comprises dNTPs at concentrations of about 1 μ M to about 100 mM, e.g., about 1 μ M to about 100 μ M, e.g., about 1 μ M to about 20 μ M, e.g., about 5 μ M to about 20 μ M, e.g., about 5 μ M to about 15 μ M, e.g., about 1 mM to about 100 mM, e.g., about 1 mM to about 20 mM, e.g., about 4 mM to about 16 mM. In some embodiments, the dNTPs within the reaction mixture are each introduced to the reaction mixture at concentrations independent of each other.

[0098] In some embodiments, user-defined data are encoded within the transitions between non-identical nucleotides within a single nucleotide polymer, producing nucleic acid data packets ("nackets"). In some embodiments, the nucleotides used to synthesize the nackets comprise A, T, C, and G. In some embodiments, the nucleotides used to synthesize the nackets comprise A, T, C, G, and U, optionally wherein the nucleotides are further modified, e.g., modified with epigenetic markers, e.g., methylation, acetylation, phosphorylation, etc. In some embodiments, one or more non-natural nucleotide may be used instead of or in addition to A, T, C, and G, and optionally U. In some embodiments, the sugar and/or backbone of the nucleotide polymer may comprise modifications, e.g., natural and/or non-natural modifications.

[0099] In some embodiments, data is encoded within the transitions between non-identical nucleotides such that the available "bits" are always one less than the number of nucleotides available to encode said data. For example, using the canonical nucleotides A, T, C, and G as the nucleotides encoding the nackets, the four nucleotides available allow for three possible transitions

from one nucleotide to the next, which yields a ternary system, i.e., “trits”. For example, if only 3 nucleotides are used to encode the nackets, the three nucleotides available allow for only two possible transitions from one nucleotide to the next, which yields a binary system, i.e., “bits”. For a further example, if five nucleotide species are used to encode the nackets, the five nucleotides available allow for four possible transitions from one nucleotide to the next, which yields a quaternary system, i.e., “quits”. In some embodiments, the nackets are encoded using three or more nucleotide species, e.g., four nucleotide species, e.g., five nucleotide species, e.g., six nucleotide species. In some embodiments, the nackets are encoded using four nucleotide species.

[0100] In some embodiments, to convert user-defined data into a population of nucleotide polymers, e.g., DNA, information is mapped to a template sequence comprising the encoding space corresponding to the number of nucleotide species used in the synthesis. For example, if using the four canonical DNA nucleotides, the user-defined data is mapped to a “trit”-based template sequence. To begin encoding data using such a trit-based template sequence, a ternary schema is first developed, e.g., the schema depicted in Figure 41. One of skill in the art will recognize such a schema is a single example of the available encoding space, and that the schema shown herein should not be construed as a limiting example. Using such a schema, a data string may be encoded from trits into DNA nucleotide transitions. For example, if the data string to be encoded comprises, e.g., 10211201, then the corresponding transitions between non-identical nucleotides would be represented by the nucleotide sequence CTGTCTATC, wherein the ternary schema of Figure 41 is used to encode the data string 10211201. (Nucleotide sequences are presented as 5' → 3' unless otherwise indicated.) However, one of skill in the art will recognize that such a nucleotide sequence is selected, in part, by the 3' terminus of the DNA strand(s) available in the reaction mixture. For example, if the DNA strand 3' terminus available for reaction is not C, as shown above, but is rather A, then the nucleotide sequence AGCGAGTGA would encode the data string 10211201, using the same ternary schema as shown in Figure 41. Thus, one of skill in the art will appreciate that it is the transitions between non-identical nucleotides that encode the user-defined data string rather than the nucleotide sequence *per se*.

[0101] Furthermore, if a non-palindromic data string is encoded into the nucleotide sequence, decoding the complimentary strand of the directly encoded nucleotide sequence may result in a reversed data string. For example, the data string 10221201 may be directly encoded into the transitions between non-identical nucleotides of sequence 5'-CTGTAGTGA-3', using the ternary

schema of Figure 41. The complimentary sequence of this directly encoded nucleotide sequence would be 3'-GACATCACT-5', which may be re-oriented as 5'-TCACTACAG-3'. Decoding the complimentary sequence 5'-TCACTACAG-3' using the encoding schema would provide data string 10212201, which is the reversed form of the originally encoded data string 10221201. In some embodiments, the reversed data string is identified by comparison to a database, e.g., a database of data strings, e.g., a database of object identification codes. In some embodiments, the encoded data strings comprise orientation sequences, which provide a sequence of encoded data that assist in identifying the proper orientation of the encoded data string. In further embodiments, the nucleotide sequence directly encoding a data string, and/or the nucleotide sequence complimentary thereto, comprises one or more nucleotide sequences and/or identifying modifications which physically and/or chemically label the nucleotide sequence and assist in identifying the proper orientation of the encoded data string.

[0102] This disclosure is directed, in part, to the synthesis of DNA sequences encoding data useful in the authentication of objects for protection against counterfeiting. This method involves first synthesizing one or more DNA sequences, incorporating said DNA sequences into an object, extracting said DNA sequences from the object when necessary for authentication purposes, and analyzing the DNA sequences for confirmation of object authenticity and/or object provenance. By encoding identification codes into DNA sequences, a highly entropic encryption system, i.e., a large permutation space, is made available for object identification. For example, if a heterogeneous population of DNA cassettes with four distinct oligonucleotide sequences are used to encode a single bit of data, and 150 rounds of cassette addition is completed as such, with each round employing a different group of four distinct oligonucleotide sequences, then 4^{150} , i.e., 2×10^{90} , different permutations of DNA sequences are synthesized, each encoding the same object identification code. This process allows access to such a large permutation space that counterfeiting by chance or estimation is effectively eliminated. Additionally, acquisition of a DNA sequence from an object followed by amplification of said DNA sequence in an attempt to include in a counterfeit product would introduce amplification biases inherent in DNA replication methods; such biases would be readily identifiable in further analysis of potential counterfeit objects.

[0103] Thus, this disclosure provides methods of confirming object authenticity and/or provenance through incorporation of DNA sequences that may be later extracted from the object and identified.

[0104] DNA is a relatively stable molecule and can be readily incorporated into or associated with goods for purposes of identifying and authenticating the goods. In certain embodiments, the nackets are adsorbed to silica beads or particles, which are optionally coated with polymer, and incorporated into goods, e.g., for purposes of identification and authentication of the goods. For example, the DNA nackets can be incorporated into silica beads, e.g., using methods as described in Koch J, et al., “A *DNA-of-things storage architecture to create materials with embedded memory*.” Nat. Biotechnol. (2020)38(1):39-43, the contents of which are incorporated herein by reference.

[0105] In certain embodiments, the nackets are incorporated into an object by direct surface conjugation. In alternative embodiments, the nackets are encapsulated into micro-containers or molecular assemblies. In certain embodiments, these encapsulated DNA sequences are incorporated into constituent parts or materials used in the production of an object, such as textiles, fabrics, leather, biomaterial products, polymers, plastics, wood, metals, inks, paints, solutions, suspensions, and raw materials. In certain embodiments, the nackets are inserted into a cell or cells, or inserted into a larger DNA construct and/or genome, such as into yeast, bacteria, fungi, plant, or animal cells, for example wherein the cells are used in the production of foods, drinks, biologics, or materials, e.g., cheese, beer, wine, vegan leather, pharmaceuticals.

[0106] In certain embodiments, the nackets, optionally incorporated (e.g., adsorbed and/or encapsulated) into beads, e.g., silica beads, are embedded into, stuck onto, or mixed into any physical material. For example, sprayer onto minerals, ores, or intermediate raw materials; embedded into polymeric thin films and used in the manufacture of any device or product; embedded into adhesives and used in the manufacture or labeling of a product; embedded into inks, e.g., used in stamping, writing, printing, inkjet printing, screen printing, or otherwise transferred to another substrate; embedded into perfume; embedded into inks used by notaries for signing documents; embedded into currency paper and/or inks; embedded into packaging for wine, spirits, and/or food; embedded into food items themselves (e.g., wine, cheese, spirits); embedded into animals used to track and trace their origin for either commercial or bioconservation reasons; embedded, sprayed, or applied to lumber products to track source and origin of lumber products;

sprayed onto or integrated into seeds for tracing seed origin/authenticity; embedded into pharmaceuticals and/or printed onto pharmaceuticals for authenticity, drug typing, identification, track and trace, and/or embedded certifications; embedded into aerospace parts for track and trace; embedded into lock-tite or equivalent thread locker to identify authenticity, part number, who applied the materials, and/or when the materials are applied. One of skill in the art would readily recognize myriad additional and/or alternative applications.

[0107] In certain embodiments, nackets incorporated into an object are extracted from the object; this extraction may be completed prior to or following production of the object, shipping of the object, sale of the object, offer for sale of the object, importation of the object, or exportation of the object. In certain embodiments, this extraction is completed for identification, authentication, and/or valuation of the object.

[0108] In certain embodiments, the nackets incorporated into an object is extracted from the object through physical and/or chemical means, such as cutting, grinding, scoring, chipping, shredding, pulverizing, dissolving, or cleaving the nackets from one or more pieces of the object.

[0109] In certain embodiments, nackets extracted from an object are isolated and/or purified; this may be accomplished by chromatography, electrophoresis, centrifugation, or combinations thereof.

[0110] In certain embodiments, nackets extracted from an object are analyzed using mass spectrometry and/or high-throughput DNA sequencing. In certain embodiments, the analyzed DNA sequences are compared to a database of object identification codes, wherein matching an object identification code to an extracted DNA sequence confirms the identity, authenticity, provenance, and/or security of the object. In certain embodiments, an analysis of the DNA sequences may be compared with results from a previous analysis of the DNA sequences from the same or similar object.

[0111] In certain embodiments, analysis of the nackets yields a “fingerprint”, wherein the specific DNA sequence, the specific cassette sequence, the sequence of transitions between non-identical nucleotides, the incidence rate of each individual nucleotide and/or cassette, the relative incidence rates of nucleotides and/or cassettes, and/or the specific molecular mass of the DNA sequence and/or its degradation products may be compared with a database of object identification codes.

[0112] In certain embodiments, the nackets are analyzed to identify the specific nucleotide sequence of said nackets, such that the identified nucleotide sequence may be used in conjunction

with the original encoding schema to decode the original encoded data string. For example, nackets comprising a series of transitions of non-identical nucleotide homopolymer extensions may be sequenced. Such nackets, e.g., synthesized using the methods above, may be variable in total length, and comprise variable lengths of homopolymer extensions. However, following sequencing of the nackets, the nacket nucleotide sequences may be compressed wherein each homopolymer extension is represented as a single nucleotide corresponding to the identity of the nucleotide comprising said homopolymer extensions. For example, continuing the example from the synthesis discussion above, nacket sequences, e.g., CCCCCCTTGGGGGGGGGGTTTTCCCTTTTTTTTAAAAAAATTTTTTCC and/or AAAAGGGCCCGGGAAAAGGGGTTTTTGGGGGGGGAAAAAA would be simplified to the compressed representative sequences CTGTCTATC and AGCGAGTGA, respectively. Continuing this example, if the exemplary schema from Figure 41 is known, then the compressed representative sequences may be decoded into the original data string 10211201.

[0113] In certain embodiments, one or more nacket may comprise a synthesis error, e.g., one or more mismatched nucleotide, one or more inserted nucleotide, one or more missing nucleotide, or a combination thereof. In certain embodiments, a population of two or more nackets are sequenced and analyzed. In certain embodiments, the population of two or more nackets are sequenced, simplified into compressed representative sequences, and then analyzed in silico. In certain embodiments, the compressed representative sequences are sorted by length of the compressed representative sequences, e.g., wherein the longest sequence(s) are “perfect” when the longest sequence(s) matches the originally encoded template sequence, and are subsequently decoded to yield the original data string. Alternatively, or additionally, the compressed representative sequences may be sorted by abundance, wherein the most abundant compressed representative sequence is selected and analyzed, optionally wherein the most abundant compressed representative sequence is further analyzed using statistical inference methods and/or models, e.g., the introduction of synchronization nucleotides, Levenshtein edit distances, maximum a posteriori estimation, Markov modeling, or a combination thereof, e.g., as discussed in Lee, H.H., *et al.*, “*Terminator-free template-independent enzymatic DNA synthesis for digital information storage.*” Nat. Commun. (2019)10:2383, the contents of which are incorporated herein by reference.

[0114] The disclosure provides methods of confirming object authenticity and/or provenance through incorporation of DNA sequences that may be later extracted from the object and identified.

[0115] In one aspect, the disclosure thus provides a method of object authentication comprising:

- i. synthesizing nackets having heterologous sequences but encoding the same data in a machine-readable code (e.g., binary or ternary code);
- ii. incorporating said nackets into or onto an object;
- iii. extracting said nackets from the object; and
- iv. analyzing the extracted nackets;
- v. optionally, comparing the analyzed nackets to a database of DNA sequences or authentication database or cryptographically hashed values;
- vi. optionally, confirming object authenticity.

[0116] In certain embodiments, the cassettes used to synthesize the nackets in the foregoing method are DNA oligonucleotide sequences comprising a 5'-overhang of one or more nucleotides, a region encoding data for identification codes, a region of complementarity to an adjacent cassette on one or both sides of the present cassette, a topoisomerase recognition sequence, and/or a 3'-overhang of one or more nucleotides. In certain embodiments, the region encoding data for identification codes comprises one or more bits of data, optionally two or more bits of data, optionally three or more bits of data, optionally five or more bits of data. In further embodiments, the region encoding data for identification codes comprises one or more bytes of data, optionally two or more bytes of data, optionally three or more bytes of data.

[0117] In certain embodiments, the cassettes are conjugated together using ligase enzymes. In alternative embodiments, the cassettes are conjugated together using topoisomerase enzymes, optionally wherein the topoisomerase is a Type I topoisomerase, such as Type IA, Type IB, Type IC, or combinations thereof, optionally wherein the topoisomerase is a Type II topoisomerase, such as Type IIA, Type IIB, or combinations thereof.

[0118] Thus, in an aspect, the disclosure provides the foregoing method of object authentication wherein in the step of synthesizing nackets having heterologous sequences but encoding the same data in a machine-readable code (e.g., binary or ternary code), the nackets are synthesized by a process comprising a series of topoisomerase-mediated ligation steps, wherein in each step, heterologous cassettes having at least two different sequences but all encoding the same data in a

machine-readable code (e.g., binary or ternary code) are ligated to a population of DNA strands by topoisomerase-mediated ligation, to provide the nackets having heterologous sequences but encoding the same data in a machine-readable code, wherein the nackets comprise a series of heterologous topoisomerase-ligated cassettes.

[0119] In another aspect, the disclosure provides the foregoing method of object authentication wherein the nackets having heterologous sequences but encoding the same data in a machine-readable code (e.g., binary or ternary code) are synthesized using a transferase-based synthesis and data encoding.

[0120] In certain embodiments, one or more DNA sequences are synthesized to encode data designed as an identification code for the object. In certain embodiments, this identification code is written manually. In alternative embodiments, this identification code is a randomly generated number or numbers.

[0121] In certain embodiments, the nackets are synthesized from a connection point on a surface, or are synthesized in solution. In certain embodiments, the nackets are synthesized in well plates, droplets, or chambers, wherein each well/droplet/chamber is used to synthesize a unique DNA sequence or sequences, wherein the DNA has a unique sequence profile but retains the data (e.g. binary code, or e.g., ternary code) encoded in the nacket. In certain embodiments, the nackets are amplified and/or replicated, optionally wherein amplification bias is used to further make the collection of DNA sequences unique. In alternative embodiments, the one or more DNA sequences are not amplified and/or replicated, and thus are directly used in incorporation into an object.

[0122] The molecules produced using topogation of heterologous components, e.g., molecules produced during surface-conjugated topogation, results in a multitude of unique molecules (if the permutation space is sufficient large). Thus, no two production runs of the exact reagents, program, and data will yield the same population of molecules. However, for useful authentication, the population of molecules produced must be known and safely stored for later authentication. To do this, an aliquot of the population of molecules produced is isolated and amplified using nucleic acid techniques, such as PCR, LAMP, isothermal amplification, and/or RCA. The result of said amplification is a solution that contains many replicates of the original unique molecules produced in the nacket. This allows for marking of very large number of objects and/or a very large area of material while maintaining a consistent, complex fingerprint throughout. A significant amount of data may be embedded into objects using such methods as described herein, through a multitude

of molecules. However, such an approach may be vulnerable to an amplification cloning attack, wherein a counterfeiter may sample the DNA embedded within an authentic object, amplify said sample, and then embed their counterfeit copies into a counterfeit object. The methods described herein are protected from this approach through the complexity of the sample; however, a further level of security may be deployed when amplification attacks are of concern.

[0123] Preventing of amplification attacks may be accomplished by writing a single nacket over a very large surface area. This single nacket is determined either from the authentication database (i.e., which specific data one is looking for) or by reference from the data file embedded in the object that is decoded from the amplified segment of the data. In one embodiment, one may write an NFT to DNA, amplify it to large volumes, and embed the resultant nackets into an object. The HASH value, a CRC, or other hash function may be computed and a molecule that is of a different length than the original (even if very close in length) would then be written over a very large area. This hash molecule would not be amplified, such that there are no replicates of the hash molecules. These hash molecules are then applied to the object as a second step, or is applied covertly to only select area(s) that should be sampled so that there is amplified material throughout the object but only a covert specific area contains the hash molecules. Alternatively, the hash molecules are mixed and embedded with the original amplified material but at low abundance, e.g., at 0.01%, 0.1%, or 1%. During authentication, the ID is read and authenticated. Next, the authentication program computes the file hash and then searches for matching sequences for the hash or other unique data string. The actual sequences of the molecules found should never be repeated. If enough sequencing has been completed and enough unique hash molecules have been found, any duplicate hash molecules indicate that an amplification attack has occurred. The molecules are nearly indiscernible from the correct molecules from a sequence and molecular perspective; thus, traditional molecular biology methods would be unable to filter or parse the hash molecules separately.

[0124] An authentication database may be used to validate sequences; however, there are several considerations in the design of the authentication that can be mitigated through information system design. The concerns are:

- Privacy: Users may not want public and/or privately posted authentication databases to have clear data corresponding to their NFTs and/or objects. The object itself could have a

serial number and/or a unique fingerprint that allows the object to be validated in a public ledger, but that has no way to identify which objects are in said ledger.

- **DNA Sequence Attack Security:** An authentication database that contains actual sequences is vulnerable to attack just like unsecure password tables. Modern IT systems have moved away from storing free text passwords and to hash tables for password management to prevent the release of user's passwords. This poses two potential risks: 1) that the free text sequences could be used to create counterfeit sequence files for authentication and sent electronically by an end point or a "man-in-the-middle" attack, or 2) that those sequences may be used to synthesize molecules to create counterfeit molecules. The authentication fingerprints are secured by not storing the actual sequences, but by storing hashes of those sequences, much like how passwords are stored in many digital systems. Further, by using hashing algorithms with salt on a by-record basis, these tables become resistant to lookup-based attacks where an attacker obtains the salt and/or the hash algorithm and then uses a brute force method to compute the hash for all possible sequences, then reverse looks up to crack the database. By having a unique salt for every record, this database becomes resistant to a reverse lookup attack. In the authentication database, the "username" or lookup value is a hash of the object's data. Within the data embedded within an object may be a segment of random data, called the object salt, that ensures that no two files will ever have the same signature. This is a block of digital information at the file layer that is created randomly at manufacturing time, is encoded into the data layer of the object, and is not retained in manufacturing logs, the authentication database, or anywhere else. In certain embodiments, digital information exists only in the object after manufacturing records are expunged. Thus, this ensures that only the bearer of the object can check its authenticity. The authenticity hashes are calculated using the full object data (across all objects) and the nacket ID of the strand that is being checked. This results in a unique salt per nacket written for which objects may have many nackets and, thus, a multitude of entries.
- **Amplification Attacks:** Another table may be maintained in the authentication database, which contains a "used unique read" table. This is calculated using the hash of the object only and no nacket id, as there is no nacket id. In this table, if a new authentication comes in requesting authorization of a molecule that was already found it may be used to invalidate the authentication request and/or warn about the collision. Here, this prevents

playback attacks and ensures only the first authentication request is approved based on a provided sequence file.

[0125] In certain embodiments, multiple packets encoding unique, distinct codes may be placed into a single object. This serves as a form of molecular encryption as one must know the encoding scheme to decode. For example, one could write hundreds of unique IDs all using different encoding schemes, e.g., different lengths, different starting sequences, and/or different ending sequences. Thus, decoding such packets requires the decoder to have previous knowledge of which encoding scheme has been used. This approach mimics a zero-trust security system. Further, one could be required to furnish a list of encoding schemes and packet IDs in a specific order. This information is the “key” to obtain a specific set of information from the object. The strength of this encoding relies on the number of unique entries and what order those things need to be placed to decode the file (or key) of interest. This approach is very powerful and functions similarly to a zero-trust security system.

[0126] In further embodiments, multiple encoding schemes may be used to read one or more code from a given sample. For example, when one is operating an authentication process that involves multiple elements on the chain of trust, each element may have its own unique code and encoding scheme. This would enable one to read, for example, the unique code (and fingerprint) for the sampling kit, the unique code (and fingerprint) for the amplification kit, and the unique code (and fingerprint) from the object of interest. When combined together, this information may be used to ensure a given combination of kits and objects may occur only once. This further strengthens the authentication process against replay attacks, man-in-the-middle attacks, and/or counterfeit authentication testing reagents.

[0127] In certain embodiments, one or more cassettes are synthesized on a chip, e.g., a chip comprising a plurality of wells and/or connection points on a surface. For example, the chip may comprise a plurality of wells and/or connection points on a surface which allow for synthesis of a plurality of heterologous sequences corresponding to one or more information sequence, e.g., a plurality of sequences, e.g., heterologous sequences, corresponding to “0”. A similar chip may allow for synthesis of a plurality of sequences, e.g., heterologous sequences, corresponding to “1”. In certain embodiments, the cassettes synthesized on a chip comprise replication/amplification primer regions, e.g., PCR primer regions, to allow for amplification. In certain embodiments, the chip comprises replication/amplification primer regions, e.g., PCR primer regions, on the acceptor

strand before addition/synthesis of cassettes. In certain embodiments, the cassettes comprise sticky ends or terminal overhangs to facilitate ligation of cassettes. For example, a first plurality of cassettes may be synthesized on a "0" chip, and a second plurality of cassettes may be synthesized on a "1" chip, wherein all cassettes comprise independently selected terminal overhangs (wherein the independently selected terminal overhangs may be the same, similar, or unique between each cassette), and wherein a binary-code sequence is synthesized by sequential addition and ligation of cassettes from either the first plurality of cassettes ("0" cassettes) or the second plurality of cassettes ("1" cassettes).

[0128] In certain embodiments, the one or more DNA sequences are incorporated into an object by direct surface conjugation. In alternative embodiments, the one or more DNA sequences are encapsulated into micro-containers, such as microspheres, such as silica microspheres. In certain embodiments, these micro-containers are incorporated into constituent parts or materials used in the production of an object, optionally wherein the constituent parts or materials are textiles, fabrics, leather, biomaterial products, polymers, plastics, wood, metals, inks, paints, solutions, suspensions, and raw materials. In certain embodiments, the one or more DNA sequences are inserted into a cell or cells, optionally inserted into a larger DNA construct and/or genome, optionally inserted into yeast, bacteria, fungi, plant, or animal cells, optionally wherein the cells are used in the production of foods, drinks, biologics, or materials, e.g., cheese, beer, wine, vegan leather, pharmaceuticals.

[0129] In certain embodiments, one or more of the DNA sequences incorporated into an object is extracted from the object. In certain embodiments, this extraction is completed prior to or following production of the object, shipping of the object, sale of the object, offer for sale of the object, importation of the object, or exportation of the object. In certain embodiments, this extraction is completed for identification, authentication, and/or valuation of the object.

[0130] In certain embodiments, one or more of the DNA sequences incorporated into an object is extracted from the object through physical means, such as cutting, grinding, scoring, chipping, shredding, or pulverizing one or more pieces of the object. In further embodiments, one or more of the DNA sequences incorporated into an object is extracted from the object through chemical means, such as dissolving or cleaving the DNA sequences from one or more pieces of the object.

[0131] In certain embodiments, one or more of the DNA sequences extracted from an object is isolated and/or purified. In certain embodiments, one or more of the DNA sequences extracted

from an object is isolated and/or purified using chromatography, for example ion exchange chromatography, size exclusion chromatography, normal-phase or reverse-phase high-performance liquid chromatography (HPLC), affinity chromatography, e.g., antibody affinity chromatography, or combinations thereof. In certain embodiments, one or more of the DNA sequences extracted from an object is isolated and/or purified using electrophoresis, for example, polyacrylamide gel electrophoresis, two-dimensional electrophoresis, pulsed field electrophoresis, Southern blotting, or combinations thereof. In further embodiments, one or more of the DNA sequences extracted from an object is isolated and/or purified using centrifugation. In certain embodiments, one or more of the DNA sequences extracted from an object is isolated and/or purified using a combination of chromatography, electrophoresis, and/or centrifugation.

[0132] In certain embodiments, one or more of the DNA sequences extracted from an object is analyzed using mass spectrometry and/or high-throughput DNA sequencing. In certain embodiments, the analyzed DNA sequences are compared to a database of object identification codes, wherein matching an object identification code to an extracted DNA sequence confirms the identity, authenticity, provenance, and/or security of the object. In certain embodiments, an analysis of the DNA sequences may be compared with results from a previous analysis of the DNA sequences from the same or similar object.

[0133] In certain embodiments, analysis of the extracted DNA sequences yields a “fingerprint”, wherein the specific DNA sequence, the incidence rate of each individual nucleotide and/or cassette, the relative incidence rates of nucleotides and/or cassettes, and/or the specific molecular mass of the DNA sequence and/or its degradation products may be compared with a database of object identification codes. In further embodiments, the sequence of DNA cassettes may be analyzed and used to determine the object identification code. In alternative embodiments, the sequence of nucleotides within heterogeneous DNA sequences and/or cassettes may be analyzed and used to determine the object identification code.

[0134] A key feature of the DNA nackets prepared as described herein is that they have a very high heterogeneity despite encoding the same digital information, e.g., binary code information. In other words, a large number of DNA sequences may encode the same data. The large permutation space afforded by using heterologous (or heterogeneous or varied) cassettes may be represented using a Heterogeneity Index (HI):

$$HI = \frac{\# \text{ DNA Sequences}}{\# \text{ Data Packets}}$$

wherein the HI is defined as a ratio between the number of DNA sequences encoding a machine readable code or data packet and the number of machine readable codes or data packets. In a traditional approach to encoding data within DNA sequences, e.g., genomic information or DNA storage by binary code, a single code is represented by a single DNA sequence, or DNA sequences of substantial similarity to the single DNA sequence accounting for occasional silent mutations and/or single-nucleotide polymorphisms (SNPs) which do not affect the amino acid sequence of the encoded protein. For naturally occurring DNA, the number of data packets (e.g., protein amino acid sequences) over the number of DNA sequences encoding the data packets would be 1 or a little more, accounting for silent mutations or variations due to infidelity of DNA replication (which has a natural error rate of about 1 in 1000 bases). In the present disclosure, a single data packet may be represented by a plurality of synonymous heterogeneous DNA sequences. For example, using heterologous (or heterogeneous or varied) cassette data writing, encoding the binary data packet 011011 with 6 rounds of addition with 2 unique cassettes per addition step, as discussed above, there is 1 machine readable code (011011) represented by 2^6 or 64 unique sequence permutations, yielding a HI of 64, i.e., 64 different synonymous sequences for one data packet. If we have a longer sequence, or a greater number of possible cassettes, e.g., a sequence encoding 100 bits, where each bit can be encoded by any of 4 different cassettes, the HI becomes very large, on the order of 4^{100} . For perspective, 4^{100} is greater than 10^{60} , and there are about 10^{80} atoms in the universe. Thus, an HI of 4^{100} implies that every single packet molecule in a given sample (or writing spot) would likely have a different DNA sequence, despite all encoding the same data packet. By contrast using homogeneous cassette data writing, where there is roughly a 1:1 correspondence between the cassette and the bit value, the HI would be approximately 1, whether the sequence is encoding 1 or 100 bits.

[0135] One consequence of this extremely high heterogeneity is that it is virtually impossible for a counterfeiter to simply amplify, analyze, and copy the DNA signature in goods labeled in accordance with the disclosure. First, the heterogeneous cassette data writing will result in very high sequence heterogeneity. With longer sequences, no two DNA molecules will be the same, making deciphering the code much more difficult for someone without the key than would be the case for a system where every DNA strand is the same. Second, even if the counterfeiter were

able to read the data packets, despite the hurdle of detecting the code within the noise created by the highly variable sequences, the counterfeiter would not be readily able to duplicate and provide counterfeit DNA markers having the unique signature produced by the relative levels (or proportions or mixtures) of the different heterologous (or heterogeneous) cassettes. As depicted in Figs. 7-13, in each round of cassette addition, the proportions of different synonymous cassettes can be varied, e.g., 50/50, 25/75, 75/25, etc., so the varying ratios of the cassettes add additional combinatorial complexity to the final mixture. The unique fingerprint provided by the particular ratio of cassettes is virtually impossible to detect and impossible to predict or counterfeit without already knowing the sequences. The cassette usage fingerprint can be varied in different ways, e.g., in an individual batch by varying the relative amounts of the cassettes for each cassette addition step, or by using two or more different large batches (e.g. amplified then mixed at different ratios) to create a "hash" providing a unique profile for the particular DNA population used to label each item.

[0136] In one embodiment, the disclosure provides a novel population of deoxyribonucleic acid sequences encoding data useful in the authentication of objects and for protection against counterfeiting (DNA 1), comprising nucleic acid data packets ("nackets"), wherein each nacket contains a plurality of DNA molecules encoding the same data, wherein the sequences of the DNA molecules are heterogeneous. For example, the disclosure provides

- 1.1. DNA 1 prepared by heterologous cassette data writing, wherein two or more cassette sequences are provided for a single bit or combination of bits in a machine-readable code, such that all or nearly all the DNA molecules in the nacket encode the same data, but the sequences of the individual molecules exhibit extremely high variation, wherein the nackets comprise a plurality of heterologous cassettes.
- 1.2. Any foregoing DNA wherein the data is in binary code.
- 1.3. Any foregoing DNA is prepared from heterologous cassettes encoding the same bit or bits of data, wherein the percent abundance of the different cassette variants used in writing the DNA provides a unique and distinguishable feature of the DNA.
- 1.4. Any foregoing DNA wherein the heterogeneity of the sequences of the DNA molecules, expressed as a Heterogeneity Index (HI) which equals the number of synonymous sequences over the number of data packets, is greater than 10, e.g., greater than 100, e.g., greater than 1000, e.g., greater than 10,000, e.g., between 10^6 and 10^{100} .

- 1.5. Any foregoing DNA wherein the data carried by the DNA is a nonfungible token (NFT).
- 1.6. Any foregoing DNA wherein the one or more DNA sequence and/or cassette contains one or more topoisomerase recognition sequences, e.g., wherein the topoisomerase recognition sequence is 5'-CCCTT-3', 5'-TCCTT-3', 5'-CCCTG-3', or 5'-TGACT-3'.
- 1.7. Any foregoing DNA wherein the one or more topoisomerase recognition sequence encodes data, e.g., wherein 5'-CCCTT-3' encodes a "1" and/or wherein 5'-TCCTT-3' encodes a "0".
- 1.8. Any previous method, wherein the DNA comprises cassettes, each cassette comprising (i) an information domain having sequence which corresponds to one or more bits in a machine-readable code, and (ii) a topoisomerase recognition sequence, wherein the cassette is 18-25 nucleotides in length.
- 1.9. Any foregoing DNA wherein the DNA is incorporated into or associated with goods for purposes of identifying and authenticating the goods.
- 1.10. Any foregoing DNA wherein the DNA is adsorbed onto, incorporated into, or encapsulated by silica beads or particles.
- 1.11. Any foregoing DNA wherein the DNA is adsorbed onto, incorporated into, or encapsulated by silica beads or particles and embedded or incorporated into goods, e.g., for purposes of identification and authentication of the goods.

[0137] In one embodiment, the disclosure provides a population of deoxyribonucleic acid sequences encoding data useful in the authentication of objects and for protection against counterfeiting (DNA 2), comprising nucleic acid data packets ("nackets"), wherein each nacket contains a plurality of DNA molecules encoding the same data, wherein the sequences of the DNA molecules are synthesized using one or more transferase enzyme, e.g., terminal deoxynucleotidyl transferase (TdT). For example, the disclosure provides:

- 2.1. DNA 2, wherein the data is user-defined data, e.g., a user-defined data string.
- 2.2. Any foregoing DNA, wherein the data is computer generated, e.g., not manually-defined, e.g., randomly computer generated.
- 2.3. Any foregoing DNA, wherein the data is in ternary code.
- 2.4. Any foregoing DNA, wherein the data is encoded and/or decoded using a schema, e.g., a schema that is user-defined, e.g., a schema that is computer generated.

- 2.5. Any foregoing DNA, wherein the one or more transferase enzyme comprises terminal deoxynucleotidyl transferase (TdT).
- 2.6. Any foregoing DNA, wherein the DNA sequences comprise a DNA initiator strand or sequence.
- 2.7. Any foregoing DNA, wherein the DNA is incorporated into or associated with goods for purposes of identifying and authenticating the goods.
- 2.8. Any foregoing DNA, wherein the DNA initiator strand or sequence comprises data useful in object authentication, e.g., lot number, batch number, production number, data code, client number, etc.
- 2.9. Any foregoing DNA, wherein the DNA sequences comprise a series of homopolymer extensions, wherein each homopolymer extension is comprised of a repeating identical nucleotide and wherein each homopolymer extension is comprised of non-identical nucleotides relative to any adjacent homopolymer extension(s).
- 2.10. The foregoing DNA, wherein the homopolymer extensions comprise one or more repeating identical nucleotides, e.g., 2 or more nucleotides, e.g., 3 or more nucleotides, e.g., 4 or more nucleotides, e.g., 5 or more nucleotides, e.g., 6 or more nucleotides, e.g., 7 or more nucleotides, e.g., 8 or more nucleotides, e.g., 9 or more nucleotides, e.g., 10 or more nucleotides, e.g., 15 or more nucleotides, e.g., 20 or more nucleotides, e.g., 25 or more nucleotides, e.g., 30 or more nucleotides, e.g., 35 or more nucleotides, e.g., 40 or more nucleotides, e.g., 45 or more nucleotides, e.g., 50 or more nucleotides, etc.
- 2.11. Any foregoing DNA, wherein the DNA comprises one or more canonical nucleotide, e.g., adenosine, guanosine, thymidine, and cytosine.
- 2.12. Any foregoing DNA, wherein the DNA comprises the canonical nucleotides adenosine, guanosine, thymidine, and cytosine.
- 2.13. Any foregoing DNA, wherein the DNA comprises one or more non-natural or non-canonical nucleotide.
- 2.14. Any foregoing DNA, wherein the DNA comprises further modifications, e.g., polyadenylation, e.g., conjugation onto small molecule and/or polymer moieties.
- 2.15. Any foregoing DNA, wherein the DNA is single-stranded.
- 2.16. Any foregoing DNA, wherein the DNA is double-stranded.
- 2.17. Any foregoing DNA, wherein the DNA is linear.

- 2.18. Any foregoing DNA, wherein the DNA is cyclic and/or cyclized.
- 2.19. Any foregoing DNA wherein the data carried by the DNA is a nonfungible token (NFT).
- 2.20. Any foregoing DNA wherein the DNA is incorporated into or associated with goods for purposes of identifying and authenticating the goods.
- 2.21. Any foregoing DNA wherein the DNA is adsorbed onto, incorporated into, or encapsulated by silica beads or particles.
- 2.22. Any foregoing DNA wherein the DNA is adsorbed onto, incorporated into, or encapsulated by silica beads or particles and embedded or incorporated into goods, e.g., for purposes of identification and authentication of the goods.

[0138] In one embodiment, the disclosure provides an ink comprising a DNA population according to any of DNA 1, et seq., and/or DNA 2, et seq. (for example a water-based ink, optionally comprising one or more pigments (for example carbon black or other pigment), binders (for example a polymer, oil, or resin), solvents (water and optionally an alcohol or organic solvent) and/or additives (e.g. drying or chelating agents)) comprising a DNA population according to any of DNA 1, et seq., and/or DNA 2, et seq. Such an ink, for example, can be used to authenticate signatures, documents or prints. In certain embodiments, a DNA population according to any of DNA 1, et seq., and/or DNA 2, et seq., in the ink encodes a non-fungible token (NFT) linked to a blockchain. Preliminary experiments suggest that the DNA will survive well in ink and paper. DNA stored on FTA cards and even dried blood spots collected on Guthrie filter cards permit accurate analysis after many years of storage without special precautions.

[0139] In one embodiment, the disclosure provides a polymer, e.g., a plastic token or object, comprising a DNA population according to any of DNA 1, et seq., and/or DNA 2, et seq.

[0140] In another aspect, the disclosure provides a method of synthesizing DNA, e.g., any of DNA 1, et seq., by topoisomerase-mediated ligation, wherein the DNA comprises cassettes corresponding to a series of bits in a machine-readable code, e.g., a binary or ternary code, comprising adding cassettes to a DNA strand, selected from

a first pool of cassettes wherein the cassettes all encode the first bit or bits, but are a mixture of at least two different sequences, and

a second pool of cassettes wherein the cassettes all encode a second bit or bits, and either all have the same sequence or are a mixture of at least two different sequences, until the desired bit sequence is reached, e.g., thereby providing a population of DNA molecules of highly heterogeneous nucleotide sequence, but all providing the same data sequence.

[0141] In another aspect, the disclosure is directed to methods of marking, identifying and authenticating goods, for example (i) methods marking the goods by incorporating or associating the DNA comprising nackets as described herein, e.g., any of DNA 1, et seq. and/or DNA 2, et seq., with the goods to be identified or authenticated, and (2) methods of identifying and optionally authenticating the goods thus marked by retrieving and sequencing the nackets, identifying the goods based on the data, e.g., binary code data, encrypted in the nackets thus retrieved and sequenced, and optionally authenticating the goods by measuring the relative amounts of the different cassette variants in the nackets thus retrieved and sequenced.

[0142] The disclosure thus provides a method of object authentication (Method 1), comprising:

- i. synthesizing DNA sequences comprising nucleic acid data packets (“nackets”), wherein each nacket contains a plurality of DNA molecules encoding the same data, wherein the sequences of the DNA molecules are heterogeneous;
- ii. incorporating said DNA sequences into or onto an object;
- iii. extracting said DNA sequences from the object; and
- iv. analyzing the extracted DNA sequences;
- v. optionally, comparing the analyzed DNA sequences to a database of DNA sequences;
- vi. optionally, confirming object authenticity.

[0143] For example, in particular embodiments the disclosure provides:

- 1.1. Method 1, wherein the DNA sequence encodes data that provides an identification code for the object.
- 1.2. Method 1.1, wherein the data that provides an identification code is randomly generated.

- 1.3. Any previous method, wherein the DNA sequences comprise any of DNA 1, et seq.
- 1.4. Any previous method wherein the DNA sequences comprise any of DNA 2, et seq.
- 1.5. Any previous method, wherein the DNA sequences is synthesized by sequential addition of one or more cassette, wherein each cassette comprises multiple nucleotides.
- 1.6. Method 1.5, wherein the cassettes are conjugated together using a ligase enzyme.
- 1.7. Method 1.5, wherein the cassettes are conjugated together using a topoisomerase enzyme.
- 1.8. Method 1.5, 1.6, or 1.7 wherein the cassettes are heterologous cassettes having at least two heterologous sequences but all encoding the same data in a machine-readable code (e.g., binary or ternary code).
- 1.9. Any foregoing method wherein the DNA sequences are synthesized by sequential addition of DNA cassettes to DNA receptor strands, wherein in each sequential addition step the cassettes comprise a heterologous population of synonymous cassettes, such that the cassettes have at least two different sequences encoding the same data in a machine-readable code (e.g., binary or ternary code).
- 1.10. Any previous method, wherein the DNA sequences are synthesized using a transferase-based synthesis and data encoding.
- 1.11. Any previous method, wherein the conjugation of DNA cassettes involves addition of a heterogeneous population of DNA cassettes, wherein each DNA cassette encodes the same one or more bits or bytes of data within distinct DNA oligonucleotide sequences.
- 1.12. Any previous method, wherein the DNA sequence is incorporated into an object by direct surface conjugation of the one or more DNA sequence onto the object.
- 1.13. Any previous method, wherein the DNA sequences are incorporated into a constituent part or material of an object used in production of said object, optionally into textiles, fabrics, leather, biomaterial products, polymers, plastics, wood, metals, inks, paints, solutions, suspensions, and raw materials.

- 1.14. Any previous method, wherein the DNA sequences are encapsulated into a micro-container, optionally a microsphere, optionally a silica microsphere, prior to incorporation into the object.
- 1.15. Any previous method, wherein the DNA sequences are encapsulated into a molecular assembly, such as a lipid nanoparticle, protein complex or aggregate, or crystal lattice.
- 1.16. Any previous method, wherein the DNA sequences are inserted into a cell or cells, optionally inserted into a larger DNA construct and/or genome, optionally inserted into yeast, bacteria, fungi, plant, or animal cells, optionally wherein the cells are used in the production of foods, drinks, biologics, or materials, e.g., cheese, beer, wine, vegan leather, pharmaceuticals.
- 1.17. Any previous method, wherein the incorporated DNA sequences are extracted from the object through physical means, optionally cutting, grinding, scoring, chipping, shredding, or pulverizing one or more pieces of the object.
- 1.18. Any previous method, wherein the incorporated DNA sequences are extracted from the object through chemical means, optionally dissolving or cleaving the DNA sequence(s) and/or one or more pieces of the object.
- 1.19. Any previous method, wherein the extracted DNA sequences are isolated and/or purified, optionally by chromatography, ion exchange chromatography, size exclusion chromatography, normal-phase or reverse-phase high-performance liquid chromatography (HPLC), antibody affinity chromatography, or combinations thereof.
- 1.20. Any previous method, wherein the extracted DNA sequences are isolated and/or purified, optionally by electrophoresis, polyacrylamide gel electrophoresis, two-dimensional electrophoresis, pulsed field electrophoresis, Southern blotting, or combinations thereof.
- 1.21. Any previous method, wherein the extracted DNA sequences are isolated and/or purified, optionally by centrifugation.
- 1.22. Any previous method, wherein the extracted DNA sequences are analyzed using mass spectrometry and/or high-throughput DNA sequencing.

- 1.23. Any previous method, wherein the extracted DNA sequences are compared to a database containing the object identification codes as originally synthesized for said object.
- 1.24. Any previous method, wherein the extracted DNA sequences are compared to results from one or more previous analysis of extracted DNA sequences from the same or similar object.
- 1.25. Any previous method, for use in combination with any of the methods of Methods 2, et seq., Methods 3, et seq., Methods 4, et seq., Methods 5, et seq., Methods 6, et seq., and/or Methods 7, et seq.

[0144] Figure 19 is a diagram showing topo cassettes (i.e., cassettes amenable to topoisomerase binding and/or topoisomerase-mediated conjugation) representing various combinations of binary bits, in accordance with embodiments of the present disclosure. Topo cassette-based chemistry is particularly well suited for data storage. Each topo cassette can be of varying length as depicted in the dashed box section with bases marked “N”. Not only can the topo cassettes be of varying length L, and they can also be of varying composition, e.g., DNA bases or other bases. Regardless of length or composition, each topo cassette can represent a single bit, two bits, four bits or 8 bits providing broad flexibility in codec development. Any number of bits per cassette may be used. However, the larger the number of bits represented, the less total number of available heterogeneous cassettes can represent a given bit pattern.

[0145] Figure 20 is a diagram showing the number of potential topo cassettes based on the number of positions and number of different DNA bases, in accordance with embodiments of the present disclosure. In particular, each position in a cassette can be represented by any of four (4) different DNA bases (G,C,A,T). The number of potential cassettes is equal to 4^N , where N equals the number of positions (or base pairs) in a cassette. For example, regarding the data-encoding portion(s) of a cassette (discussed here as distinct from the topoisomerase recognition portion and/or overhang portion, though these regions may encode further information), a 10 base pair (or position) cassette would have a number of potential cassettes = 4^{10} or 1,048,576 unique cassettes. Other example sizes and number of potential cassettes are shown, such as: $4^{20} = 1,099,511,627,776$ (20 position cassette); $4^{19} = 274,877,906,944$ (19 position cassette); $4^{18} = 68,719,476,736$ (18 position cassette), and the like. This cassette flexibility of cassette length and

composition provides a nearly infinite palette of cassettes for use in data writing. If a cassette has ten positions for example, any single position could be any of the 4 chemical bases of DNA, as shown at the top of Figure 20. Hence a 10 position Topo cassette can have over 1 million potential unique cassettes. In some embodiments, the topo cassettes may range in size between 18-20 base pairs (bp). The potential palette of cassettes is illustrated in the Figure 20. A 20 bp cassette size would enable ~1.1 trillion potential unique cassettes to choose from. The permutation space of possible cassettes would increase exponentially further with the use of additional synthetic bases, such as Q and R, which when added to the 4 bases, the combinations would jump to 6^{20} vs 4^{20} . [0146] Figure 21 is a diagram showing how multiple different (or unique) cassettes may be used to specify the same underlying binary information, in accordance with embodiments of the present disclosure. In this regard, topo cassettes may be used to make replication resistant or attack resistant, encrypted molecular tags or codes by creating multiple cassettes that specify the same 2-bit binary code. Billions of Topo cassettes representing the same binary information can be constructed. Also, substituting any single base will change the underlying binary code represented and damage to single bases changes the binary code represented. For example, for a 10-cassette memory string (or nacket), the number of possible molecular structures = $4^{10} = 1,048,576$, that can represent the same 20-bit binary code, with 4 unique cassettes per 2-bit binary code. Figure 21 also shows the starter strand (or starter string) (SS) or acceptor strand of DNA that is attached to a substrate on one end and an end cap (EC) DNA strand at the end of the DNA string or Nacket, with a plurality of data cassettes, which may be topo cassettes, between the SS and the EC.

[0147] The writing of digital data in synthetic DNA may be thought of in terms of single base synthesis and two-bit per base encoding where any of the four DNA bases can represent the two-bit combinations of 00, 01, 10, and 11. With such an encoding scheme, substituting any single base with another accidentally during synthesis changes the fundamental binary code being represented. A similar situation occurs when any given single base is damaged. Thus, such a scheme is not desirable for data storage or secure code generation or authentication.

[0148] In one embodiment of the topo cassettes described herein, each set of two bits is represented by a multi-base, double stranded DNA cassette. In this case, damage to any single base would not prevent the accurate reading of the underlying binary, especially in the instance of longer cassettes, based on error checking and error correction. Furthermore, in the instance of a 20bp cassette system there would be ~1.1 trillion (4^{20}) theoretical topo cassettes that could be equally

apportioned amongst the four, 2-bit combinations shown. If a 4-bit encoding structure was chosen, it would be ~ 1.1 trillion/ $(2^4 = 16)$ four-bit permutations. It is also possible for each combination of bits to be represented by topo cassettes of differing sizes – e.g. 00=20bp cassettes, 01=18 bp cassettes, 10=16 bp cassettes, 11=19 bp cassettes. Any other cassette sizes may be used.

[0149] Figure 22 is a diagram showing a comparison of homogeneous cassette data writing and heterogeneous cassette data writing using a plurality of topo cassettes combined in a predetermined formulation or mixture, in accordance with embodiments of the present disclosure. Topo cassettes can also be used to make replication resistant or attack resistant, encrypted molecular tags or coded data by combine multiple unique cassettes in varying formulations or mixtures, but the underlying binary information remains the same. For example, each two-bit combination may be represented by Y different cassettes simultaneously in specific formulations/mixtures. Thus, sequences can be formulated in varying ratios for additional combinatorial complexity, such as: $(100^Y)^4$ formulations possible, assuming integer percentages of each potential sequence in the formulation or mixture, for a 2-bit binary encoding scheme.

[0150] Accordingly, a further step possible to encrypt the underlying binary information is to write any set of binary information (e.g., 00, 01, 10, 11) not with any single topo cassette, but a multitude of cassettes, mixed in fixed ratios for each set of binary bits in a production run. An example is illustrated in the Figure 22. In the example illustrated, each set of two bits is represented by 4 different Topo cassettes, those four topo cassettes are used in mixtures/ formulations of fixed ratio. The ability to combine cassettes in mixtures or formulations further broadens the permutation space for the cryptographic writing of binary sequences and enhances authentication (discussed more hereinafter). Further, the mixture or formulation may be changed with each production run. In some embodiments, the lot number may be directly correlated to the mixture ratios used for that lot.

[0151] Figure 23 is a diagram showing the example heterogeneous mixtures/formulations of topo cassettes shown Fig. 22 loaded into print heads 830,832,834,836 of a laser jet DNA printer, in accordance with embodiments of the present disclosure. In particular, Figure 23 shows a side view of a silicon wafer 10 having a patterned (or un-patterned) layer 202 of SiO₂ on top of the Si wafer 10 to form spot pillars (or spots) 14 with an attachment top coating 204, e.g., HfO₂, and fluid channels 15 between the spots 14, and also shows a side view of a print head bank 822 having four nozzles 830A, 832A, 834A, 836A corresponding to the binary code 2-bit pairs (00,01,10,11), for

2-bit binary encoding, for adding cassettes associated with same, and a fifth nozzle 814 for writing the deblock/adapter, and also shows that a wash cycle using a wash fluid 820, may be spread, flowed, applied or sprayed horizontally across the wafer surface or vertically as a separate print head 816 and corresponding nozzle 816A as part of the print head bank 822, similar to that described in the aforementioned commonly owned patent application on inkjet printing DNA. The print head bank 822 may be controlled by a print head controller (discussed hereinafter with Fig. 30A) to move (as a group) as shown by arrow 818 across the wafer array to deliver the desired droplet at precise spot locations. In this case, the print head or print head bank 822 has four chambers 830, 832, 834, 836 with associated nozzles 830A, 832A, 834A, 836A, respectively, with reagents used to adding codes via droplets to the starter DNA strands (or starter strands or starter strings or SS) 210 in the liquid bubble 802 shown on the top of each spot 14, e.g., Add "00" head 830, Add "01" head 832, Add "10" head 834, Add "11" head 836, and Deblock/Adapter head 814. The Add 00,01,10,11 reagents may add the "cassettes" described herein comprising a plurality of double-stranded DNA bases as discussed herein, and the addition reaction chemistry functions the same as that described above and in the commonly owned US patents and patent applications.

[0152] More specifically, each of the chambers has a predetermined mixture 830B, 832B, 834B, 836B, of a plurality of cassettes C1-C16, associated with each 2-bit pair, e.g., C1-C4 corresponds to "00" bits, C5-C8 corresponds to "01" bits, C9-C12 corresponds to "10" bits, and C13-C16 corresponds to "11" bits, and each mixture is loaded into the corresponding chambers 830, 832, 834, 836, respectively, of the print head bank 822 before the writing process begins.

[0153] In particular, In some embodiments, the addition chemistry used for writing to the polymer may be the chemistry described herein and in the aforementioned commonly-owned US patents, which comprises a "deblock" step. Also, in some embodiments, the addition chemistry used for writing to the polymer may be the chemistry described in the aforementioned commonly owned pending US patent applications where an "adapter" is used instead of a deblock enzyme. Accordingly, the action of getting the DNA strand ready to perform another addition reaction, may be referred to herein as a "deblock/adapter" or "adapter BA" action.

[0154] In some embodiments, a wash fluid is flowed over the array to after an addition reaction to prepare the DNA for the next addition reaction or deblock reaction. In some embodiments, instead of or in addition to having the side flow wash shown, the print head may have an additional chamber or nozzle (shown in dashed lines) that has a wash fluid in it that is dispensed during the

wash cycles. Also, in some embodiments, the deblock/adapter print head may be applied or flowed across the wafer which is flowed during the appropriate times during the write process.

[0155] Figure 24 is a diagram showing a process for writing two-bit binary codes onto the surface of a substrate or matrix, in accordance with embodiments of the present disclosure. In particular, Figure 24 shows a cross-section side view of a silicon wafer 10 (as an example substrate) with patterned layers 202,204 showing starter polymer or DNA strands (SS) 210 in liquid 802 attached to spot pillars (or spots) 14 and showing a side view of a data writing (or printing) process 930 to add bits or codes to the free end of starter polymer DNA strands on the wafer, similar to that described in the aforementioned commonly owned patent application on inkjet printing DNA. In particular, a write addition begins by performing a wash cycle 820 to prepare the DNA strands 210 for the first write addition reaction. Next, the print head dispenses an Add “00”, “01”, “10”, or “11” droplet onto the desired spot location(s), the droplet comprising the cassettes or cassette mixture/formulation associated with the 2-bit code being written, as shown by blocks 912A, 912B, 912C. After the addition reaction is complete, a wash cycle 802 is performed to prepare the DNA strands for the deblock/adapter reaction. Next, the print head dispenses a Deblock/Adapter droplet onto the desired spot location(s) that have just had an addition reaction, shown by blocks 904A, 904B, 904C. After the Deblock/Adapter reaction is complete, a wash cycle 802 is performed to prepare the DNA strands for the next addition reaction. Next, the print head dispenses an Add “00”, “01”, “10”, or “11” droplet onto the desired spot location(s), depending on the desired cassette(s) or cassette mixture/formulation associated with the 2-bit code being written, as shown by blocks 916A, 916B, 916C. After the addition reaction is complete, a wash cycle 820 is performed to prepare the DNA strands for the deblock/adapter reaction. Next, the print head dispenses a Deblock/Adapter droplet onto the desired spot location(s) that have just had an addition reaction shown by blocks 908A, 908B, 908C. The above process repeats until all the desired cassettes or 2-bit codes have been written to the DNA strands. The write addition process is also discussed further with regard to Fig. 31A and 31B hereinafter.

[0156] In some embodiments, instead of the deblock/adapter steps 904A, 908A, when using the AB/BA writing approach, no AB adapter may be needed if the cassettes are designed with an AB and a BA for each binary code, such as is shown in Fig. 6, and in the code writing example of Figs. 7-13. Also, in that case, in the writing (printing) logic 3100 of Fig. 31A and 5300 of Fig. 53A, block 3120 would not be performed.

[0157] Figures 25A, 25B, 25C, 25D, 25E, 25F, 25G, 25H, 25I, and 25J are diagrams showing a process for writing memory strings at a spot on a substrate using a pre-set formulation or mixture of cassettes for each 2-bit pair, in accordance with embodiments of the present disclosure. In particular, it shows each write cycle and how the cassettes are added to the memory string. For each write, the cassettes in the droplet will randomly attach to the loose strings. In this example, 10 independent DNA chains or memory strings are being synthesized, each representing the same binary code of 20 bits shown (11010010011111001001).

[0158] If the first set of two bits is 11 as shown, with the current formulation, 50% of the molecules on the surface will get cassettes labelled C13 and 50% will get cassettes labelled C16. Which of the 10 molecules on the surface will get which of the two “11” cassettes is truly random. This non-algorithmic randomness will also ensure that data written with such a formulation (or encryption) approach will likely be quantum replication resistant or attack resistant since all algorithmic random number generators have subtle biases which quantum computers can potentially hack.

[0159] The process continues with each of the 10 DNA strands being synthesized getting a random cassette based on the formulation being used to represent the relevant two bits, which in Fig. 25B, are the bits 01. The molecular representation on the surface across all strands being synthesized will however statistically reflect the ratios of cassettes in the particular mixture/formulation for those two respective bits of binary as illustrated with the cassette labels C1-C16, each being a unique cassette. Similar process occurs for Figs. 25C-25J.

[0160] Referring to Fig. 25J, at the end of the synthesis run in this example, the molecules generated will truly be random, but they will all represent the same underlying 20-bit binary information shown above (11010010011111001001). In this example, a string of 10 cassettes was constructed. If each set of two-bit binary codes had four distinct topo cassettes used, there would be over 1 million (4^{10}) distinct molecules possible, as discussed herein above. The possible permutations of data chains or memory strings is: $(\# \text{ cassettes/bit pair})^{(\text{cassettes in a chain})}$.

[0161] If a memory string of 128 cassettes was constructed, and each set of two-bit binary codes had four distinct cassettes used, there would be over $1.16e77$ (4^{128}) distinct molecules possible. Such a permutation space would jump radically if each set of two-bit binary codes were represented by say 10 cassettes (discussed further with Figs. 29B and 29C). The resulting permutation space is so large and random, that it is economically unviable to synthesize the range

of molecules needed to fake an NFT token or smart contract or other secure data file encoded in this manner.

[0162] Figure 26 is a diagram showing a cassette along a memory string and cassettes assigned to each 2-bit code in the memory string, in accordance with embodiments of the present disclosure.

[0163] Figures 27A, 27B, 27C, and 27D are diagrams showing a process for validating memory strings or nackets using the predetermined cassette mixture associated with a given 2-bit binary code, in accordance with embodiments of the present disclosure. In particular, in Fig. 27A, the all the cassettes associated with the 11 bit code are collected and analyzed, the total distribution should approximately match the mixture or formulation for the associated lot number. Similarly, in Fig. 27B, 27C, and 27D, the cassettes associated with the 01, 00, 10, bit codes respectively are separately collected and analyzed, the total distribution for each bit code should approximately match the mixture or formulation for the associated lot number. Thus, the use of a mixture or formulation of cassettes assigned to a bit code provides another dimension of randomness and authenticity.

[0164] In particular, Figure 28 shows a diagram illustrating two dimensions of randomness and validation of memory strings (or nackets) in accordance with embodiment of the present disclosure. A first dimension is along a memory string, where validation may be based on the multiple cassette assignment for each 2-bit code for a given lot number. The second dimension is across all memory strings along the surface for each spot, where validation is based on mixture or formulation associated with the 2-bit code for a given lot number. This is also shown in the flow diagram of the decoding and mixture confirmation logic Fig. 34. Alternatively, the two dimensions of randomness may be described as a Production Fingerprint and a Molecular Fingerprint. In such a case, a Production Fingerprint comprises the underlying information encoded within a nacket, wherein the variability potential provides a high-entropy space of variability and, optionally, randomness. A Molecular Fingerprint comprises the physical molecular structure (e.g., DNA nucleotide sequence) of the nacket, which provides a distinct and orthogonal (relative to the Production Fingerprint) high-entropy space of variability and randomness, e.g., wherein each molecule encoding information may be unique.

[0165] Figures 29A, 29B, and 29C are binary code to cassette tables showing various assignments between binary codes and cassettes and associated cassette mixtures/formulations, based on lot numbers, in accordance with embodiments of the present disclosure. Also, the information in the

binary code to cassette table may be stored on and retrieved from the blockchain, as discussed herein. In particular, Fig. 29A shows a binary code to cassette table, sorted by lot number, for a 2-bit encoding scheme, having 4 cassettes assigned to each 2-bit binary code, and a predetermined mixture/formulation for each 2-bit code. In this case, Lot 1 shows cassettes C1-C16 being assigned to 2-bit codes as shown in the example described herein with Figs. 22 and 23. In that case, C1-C4 are assigned to “00” bit code, C5-C8 are assigned to “01” bit code, C9-C12 are assigned to “10” bit code, and C13-C16 are assigned to “11” bit code. Also, the proportions for the mixture % also being the same as the example described herein with Figs. 22 and 23. For Lot 2, the assignment of cassettes (C’s) was rolled or shifted 1 position down. In that case, C16 is at the top, followed by C1-C15, and the resulting 4 cassettes for each 2-bit code are assigned accordingly as shown under Lot 2 in Fig. 29A. Also, the proportions for the mixture % for each group of 4 were randomly scrambled from that shown in Lot 1. For Lot 3, the assignment of cassettes (C’s) and mixture precents (%s) were chosen at random and not related to the prior assignments. For Lot 4, the assignment of cassettes (C’s) was rolled or shifted by groups of 4 cassettes from that shown in Lot 1. In that case, C13-C16 is at the top, followed by C1-C4, C5-C8, and C9-C12, and each 2-bit code are assigned accordingly as shown under Lot 4 in Fig. 29A. Also, the proportions for the mixture % for each group of 4 were scrambled from that shown in Lot 3. The above examples of different assignments and mixture percentages (or ratios) are for illustrative purposes. Any other numbers and variations may be used.

[0166] Referring to Figure 29A, in some embodiments, the Binary Code to Cassette Table may also include a writing direction (Write Direction) to be used for a given Lot for writing the digital code, such as MSB-LSB, LSB-MSB, or Random. In particular, a memory string or nacket to be written at a given spot on the chip from the surface of the substrate or wafer array may be written in two possible directions: MSB-LSB, from most significant bit(s) (MSB) to least significant bit(s) (LSB), i.e., from left to right, or LSB-MSB, from least significant bit(s) (LSB) to most significant bit(s) (MSB), i.e., from right to left. The number of bits for the LSB or MSB will depend on the type of encoding used, as discussed hereinafter. Also, a writing direction of “Random” for a given lot indicates that the writing logic can decide which direction to write any given spot within a given lot number. In that case, the writing direction may be indicated by a flag or code (e.g., an MSB-LSB/LSB-MSB flag or code, where 1 = MSB-LSB and 0 = LSB-MSB) may be written in the end cap (EC) for all the memory strings or nackets written at a given spot. Thus, a given lot number

may have a random distribution of writing directions on the same chip or array. As a result, a plurality of spots written with the same code for redundancy and error detection/correction, may have the codes written into the memory string or nacket in one of two different directions selected randomly. This adds another level of randomness to the resulting encoded DNA/polymer memory string or nacket.

[0167] For example, the code 10101100, when written MSB-LSB (using single bit encoding), would have the LSB (0) nearest to the end cap. However, the same code, 10101100, when written LSB-MSB, would have the MSB (1) nearest to the end cap. In the case of 2-bit binary encoding, the LSB and MSB comprises two binary bits. Thus, for the code 10101100, when written MSB-LSB, would have the LSB (00) nearest to the end cap, and, the same code 10101100, when written LSB-MSB, would have the MSB (10) nearest to the end cap.

[0168] Also, while some of the examples herein show 1-bit binary codes, and some show 2-bit binary encoding, it should be understood that any number of bits may be used for encoding the binary data to cassettes, as discussed herein (e.g., Fig.19). In some embodiments, the encoding scheme may change for a given lot number, which would be saved in the Binary Code to Cassette Table. For example, Lot 1 may have 2-bit encoding, Lot 2 may have 3-bit encoding, Lot 3 may have 4-bit encoding, and the like for other lots. In that case, for 3-bit encoding, there would be eight different binary codes 000-111 where each code is assigned one or more cassettes. If there were 4 cassettes assigned per code, then such a configuration would use 32 unique cassettes. Similarly, if 4-bit encoding was used, there would be 64 codes 0000-1111, where each code is assigned one or more cassettes. If there were 4 cassettes assigned per code, then such a configuration would use 64 unique cassettes. In some embodiments, the data to be written may be padded with a predetermined number of extra bits to make the total number of bits divide evenly into the bit encoding scheme.

[0169] Fig. 29B shows a binary code to cassette table, sorted by lot number, for a 2-bit encoding scheme, having a variable number of cassettes assigned to each 2-bit binary code based on lot number, and a predetermined mixture/formulation for each 2-bit code. In this case, Lot 1 shows 4 cassettes assigned to 2-bit codes as shown in the example described herein with Figs. 22 and 23. Lot 2 shows 5 unique cassettes assigned to each 2-bit code for a total of 20 cassettes (C1-C20). Lot 3 shows 6 unique cassettes assigned to each 2-bit code for a total of 24 cassettes (C1-C24). Lot 4 shows 7 unique cassettes assigned to each 2-bit code for a total of 28 cassettes (C1-C28).

Lot N shows 10 unique cassettes assigned to each 2-bit code for a total of 40 cassettes (C1-C28). As the number of cassettes in a mixture increases, the proportions are reduced, which may be balanced against the % threshold or tolerance of the detection system to optimize validation accuracy. X indicates not applicable.

[0170] Fig. 29C shows a binary code to cassette table, sorted by lot number, for a 2-bit encoding scheme, having 4 cassettes assigned to each 2-bit binary code based on lot number chosen from a total of 40 cassettes, 10 cassettes per 2-bit binary code, and a predetermined mixture/formulation for each 2-bit code. In this case, Lot 1 shows 4 unique cassettes out of a possible 10 assigned to 2-bit codes. Lot 2 shows a different 4 cassettes assigned to each 2-bit code selected out of 10 cassettes. Lot 3 shows a different 4 cassettes assigned to each 2-bit code selected out of 10 cassettes. Lot 4 shows a different 4 cassettes assigned to each 2-bit code selected out of 10 cassettes. Lot N shows a different 4 cassettes assigned to each 2-bit code selected out of 10 cassettes. X indicates cassettes that were not used for a given lot. An advantage of the approach in Fig. 29C is there only needs to be 4 cassettes combined in the mixture for any given lot, which increases the percent proportions, while also maintaining randomness by having 10 cassettes available for any given 2-bit code. In some embodiments, each 2-bit pair may have access to all 40 cassettes when selecting the cassettes for a given 2-bit code. Also, the same approach may be used for any number of cassettes for a given 2-bit code, e.g., select 5 cassettes out of 40 cassettes. Also, the number of total cassettes may also be increased to increase randomness, if needed.

[0171] Figure 30A is a block diagram showing an inkjet printing system 1900 including an inkjet printing instrument 1902 and a computer system 1904 which interfaces with the instrument 1902, similar to that described in the aforementioned US patent application on inkjet printing with DNA. The inkjet printing instrument 1902 may include the piezo-electric inkjet print heads 1906, which deliver the reagent droplets discussed herein to the desired writing spots on the wafer array 10, which is mounted to an XY stage 1907. The print head and XY stage may be controlled by a print head and array stage controller and inspection logic 1908, which communicates with Local Control Logic 1910 to write the desired reagents and codes to the DNA strands as directed as discussed herein. For example, one or more of the read/write address and/or data inputs, outputs and/or control lines 1912, may be received from or provided to a serial bus, which includes commands for which codes or data to write to the array. The Computer System 1904 may receive commands from a user 1903 and provide information to a display 1905 for use by the user 1903, and may also

provide commands to the local control logic 1910, which provides specific write requests to a print head/bank 1906 and to array stage controller and inspection logic 1908. The print head and array stage controller and inspection logic 1908 controls the print head position XYZ and the wafer array XY stage 1907, and also receives data from a droplet viewer (or sensor) 1911 to determine quality control of the drops and reports results and errors back to the local control logic 1910 and the computer system 1908, which stores the droplet error information on a DNA Data Server 1915 or other memory device for future use when reading the data. Such information may be used to correct or ignore certain data that is known to have certain errors in the data caused by droplet errors.

[0172] The inkjet printing instrument 1902 may include instrument (fluidics/reagents) control logic 1914 which controls the reagent supplies 1916 to the print head 1906 and controls the fluid flows 1920 through a flow inlet manifold 1921, across the wafer array 10, e.g., wash fluid 1922, cleaving fluid 1924, preparation fluid 1926, and the like, via valves 1920A, 1920B, 1920C, respectively, and control lines 1919, as well as controls the exiting fluids 1930 which flows through a flow exit manifold 1931, such as the waste fluid 1932 via valve 1930A and control lines 1933, and the fluid 1934 having the coded DNA that has been detached from the wafer array 10 via valve 1930B and control lines 1933, and collected, e.g., in a collection bin 1936, for later reading. The reagents/supply loading components may be controlled by the instrument 1902 and may include the necessary known valves and fluidics to load the print head/bank with the desired cassette assignments and mixtures/formulations associated with binary codes for a given lot based on data from the Binary Code to Cassette Tables discussed herein with Figs 29A-29C, which data may be stored in the DNA Data Server 1915 or other memory device and provided to the instrument 1902 by the computer system or the local control logic 1910 or may access the server directly.

[0173] In some embodiments, the print head and array stage controller 1908 may be configured to swap out (remove/load) the print heads or print head bank (group of print heads) between each production lot writing of DNA. In that case, the print head and array stage controller 1908 may remove the existing print heads/bank 1906 and obtain the corresponding print heads/bank 3004 having the desired mixtures/ formulations (C1-Cm) and load the print heads/bank into the inkjet printer for writing the DNA/polymer to the wafer array 10. This may be performed by a robotic arm 3002 or other controllable device or system which may be part of or separate from the print head and array stage controller 1902.

[0174] Figure 30B is a block diagram of the computer system 1904 of Figure 30A, in accordance with embodiments of the present disclosure. The computer system (Fig. 30B) 1904 may interact with the inkjet printing instrument 1902, and may also interact with the instrument control 1914, which interacts with separate fluid supplies 1916 and the like, all of which interact with one or more CPU/Processors 1952 or logic for performing certain functions described herein. Also, the Computer System in Figs. 30A and 30B may interface with a user 1903 and a display screen 1905 (Fig. 30A).

[0175] The Local Control Logic 1910 and the Fluidics Instrument Control 1914 and the print head and array stage controller 1908, have the necessary electronics, computer processing power, interfaces, memory, hardware, software, firmware, logic/state machines, databases, microprocessors, communication links, displays or other visual or audio user interfaces, printing devices, and any other input/output interfaces, including sufficient fluidic and/or pneumatic control, supply and measurement capability to provide the functions or achieve the results described herein.

[0176] Figure 31A is a flow diagram 3100 for writing (printing) and unloading coded polymer memory strings in an inkjet writing system, in accordance with embodiments of the present disclosure, which logic 3100 may be performed by the system of Fig. 30A. In some embodiments, the above writing process may be repeated for each new set of DNA strings to be written. In particular, the logic 3100 begins at block 3102 by loading or printing starter (or acceptor) DNA strands (or starter strand or SS) onto the wafer array spots 14 (Fig. 23). Next, block 3104 receives the Lot # and the Binary Code to print/write the first memory string or nacket. Next, block 3106 of the logic retrieves 4 inkjet cartridges or heads for each of the 2-bit codes (00,01,10,11) each code being assigned a different mixture of DNA cassettes from the DNA Data Server for the Lot #, such as from the corresponding Binary Code to Cassette Table 2900, 2920, 2904 (Figs. 29A, 29B, 29C). Next, block 3108 of the logic retrieves the writing direction from the Binary Code to Cassette Table stored on the DNA Data Server for the given Lot# and if Random, randomly selects the writing direction for the spot or spots to be written and saves it in the Binary Code to Cassette Table. Next, block 3110 of the logic 3100 retrieves the first 2-bit binary code to be written to the substrate or wafer from the desired Binary Code, based on the writing direction obtained from the Binary Code to Cassette Table. Next, block 3112 of the logic 3100 performs a wash cycle across the wafer array to clear any extraneous reagents from the surface of the wafer. Next, block 3114

of the logic writes/prints the 2-bit code to the memory string/nacket with the appropriate cassettes at the desired spot(s) per a writing process described herein with Fig. 31B. After the 2-bit code is written, block 3116 of the logic determines whether there are more spots to be written before the Deblock/Adapter is applied to the spot. If Yes, the logic goes back to block 3114 and writes/prints the 2-bit code to the memory string/nacket with the appropriate cassettes at the desired spot(s) per a writing process described herein with Fig. 31B until all desired spots are written for that 2-bit code. Then, when the result of block 3116 is No, block 3118 waits for the addition reaction to complete. When the reaction has completed, block 3120 prints the Deblock/Adapter for the desired spots. In some embodiments, the Deblock/Adapter may be washed across the surface of the array instead of using an inkjet cartridge or head for the Deblock/Adapter (as shown in Fig. 23). Next, block 3122 determines whether all 2-bit codes have been written for the current string or nacket. If not, block 3124 gets the next 2-bit code in the string and proceeds back to block 3112 to perform the wash cycle and repeats the process for the next 2-bit code, as shown in Fig. 31A. If the result of block 3122 is Yes, all the 2-bit codes have been written, and block 3216 of the logic writes/prints the end cap onto the memory string or nacket at the appropriate spots on the wafer with writing direction information encoded into the end cap, e.g., by a flag or other means, if writing direction is used or appropriate for the given application. Other techniques for encoding or flagging the writing direction may be used if desired. Next, block 3128 determines whether all memory strings or nackets have been written for the wafer array or chip. If Not, block 3130 of the logic gets the next desired Binary Code for the next memory string or nacket to be written and proceeds back block 3110 to retrieve the first 2-bit binary code from the desired Binary Code to be written for the next string, and the logic repeats the process for writing the next desired Binary Code until all usable desired spots are written on the wafer array or chip, or all desired Binary Codes have been written. Next, when the result of block 3128 is Yes all spots or codes have been written and block 3132 washes the wafer array with cleaving fluid and unloads and captures the DNA/polymer memory strings or nackets in a containment bin (for future reading), such as that shown in Fig. 30A and Fig. 33, and the logic exits.

[0177] Figure 31B is a flow diagram for writing (printing) 2-bit code to DNA/polymer memory string in an inkjet writing system, in accordance with embodiments of the present disclosure, which logic may be performed by the system of Fig. 30A. In particular, the logic checks for each 2-bit code to be written and causes the appropriate inkjet cartridge or head having the corresponding

DNA cassette (or topo-cassette) mixture to print the appropriate 2-bit code at the desired spot(s)/location(s) on the wafer array or chip. Once the appropriate 2-bit code has been written, for the desired number of spots, the logic determines whether any droplet errors were detected by the droplet viewer, which may be part of the print head and array stage controller and inspection logic. If any errors were detected the logic saves the error location(s) and bit number for future reading, and the logic exits. In particular, the logic 3150 begins a block 3152 which determines whether the 2-bit code to be written is "00" bits. If yes, block 3154 prints the "00" bits with the "00" ink cartridge having the "00" DNA cassette (or topo-cassette) mixture at the desired spot or spots/location on the wafer array or chip. Next, or if the result of block 3152 is No, block 3156 determines whether the 2-bit code to be written is "01" bits. If yes, block 3158 prints the "01" bits with the "01" ink cartridge having the "01" DNA cassette mixture at the desired spot or spots/location on the wafer array or chip. Next, or if the result of block 3156 is No, block 3160 determines whether the 2-bit code to be written is "10" bits. If yes, block 3162 prints the "10" bits with the "10" ink cartridge having the "10" DNA cassette mixture at the desired spot or spots/location on the wafer array or chip. Next, or if the result of block 3160 is No, block 3164 determines whether the 2-bit code to be written is "11" bits. If yes, block 3166 prints the "11" bits with the "11" ink cartridge having the "11" DNA cassette mixture at the desired spot or spots/location on the wafer array or chip. Next, or if the result of block 3164 is No, block 3168 determines whether the bit writing is complete for the spot or group of spots on the chip. If Yes, block 3170 determines whether there are any droplet errors were detected by the droplet viewer (or sensor) 1911 (Fig. 30A), which may be part of the print head and array stage controller and inspection logic 1908 (Fig. 30A). If Yes, errors were detected and block 3072 saves the error location(s) and bit number for future reading, and the logic exits. If the result of block 3072 is No, then no droplet errors were found for that write cycle and the logic exits.

[0178] Figure 32A is a side view diagram showing several spots 14 with coded DNA strands 1002, 1004, 1006 (using the code writing approach described herein) and cleaving fluid 1008 for removing coded DNA strands from surface of substrate, in accordance with embodiments of the present disclosure. In particular, Figure 32A is a side view of a silicon wafer 10 with patterned layers 202, 204 (as an example substrate or wafer) showing starter (or acceptor) DNA strands 210 attached to spot pillars (or spots) 14 at one end of the starter DNA and attached to coded DNA on the other end, and also showing how a cleaving fluid 1008 may be used to remove the coded DNA

strands 1002, 1004, 1006 from the wafer 10. In some embodiments, the wafer 10 may be unpatterned or partially patterned, if desired, as discussed in the aforementioned commonly owned patent application relating to inkjet printing of DNA.. Each pillar or spot 14 has a plurality of coded polymer or DNA strands (or nackets). When all the bits or cassettes or codes have been written or printed, a cleaving fluid 1008 may be flowed across the wafer array (or chip), which releases the coded DNA 1002, 1004, 1006 allowing them to be removed or flowed (shown by an arrow 1010) from the solid substrate 204 and placed in a storage container (Fig. 33) which may contain liquid to keep the memory strings hydrated or may allow them to dehydrate for later re-hydration and reading.

[0179] Figure 32B is a diagram showing an array of spots with coded DNA having columns (X) of redundant spots with the same encoded DNA data written, and rows (Y) of spots with different encoded DNA written, in accordance with embodiments of the present disclosure. In some embodiments, each spot on the surface of the substrate or wafer may have unique encoded data written, which may include an address or ID associated with the memory strings or nackets or chains written to that spot, e.g., memory string (or nacket) address or ID, such as NID1, NID2, NID3, NID4, to NIDY. In some embodiments, the same unique encoded data may be written to a plurality of spots across the surface of the substrate or wafer to provide redundancy and increased error checking and validation. In that case, the redundancy and validation discussed herein may be performed for all the memory strings (or nackets) having the same address or ID, independent of which spots or how many spots the strings or nackets started from. This increases the number of strings that are part of the vertical and horizontal redundancy and validation discussed herein.

[0180] In particular, Figure 32B shows a plurality of spots having the same memory string or nacket ID or address. For example, there are multiple spots (shown as X spots) in the first row with the same Nacket ID, NID1, and multiple spots (X) in the second row with the same Nacket ID, NID2, and similar redundancy for subsequent rows, where the same data is written across multiple spots on the chip, which provides redundancy and fraud checking and error checking capability. In that case, all memory strings or nackets with the same Nacket ID (or memory string address) may be analyzed as a group for the validation check. While the same code may be written across multiple spots, the actual DNA/polymer sequence of bases or groups of bases (cassettes) will be different from one spot to the next or even within the same spot due to the

mixing/formulations of cassettes discussed herein and due to the writing direction discussed herein.

[0181] Figure 33 is a diagram showing removal of spotted DNA memory strings or nackets from the surface of substrate to a collection bin and reading and decoding the DNA collection, in accordance with embodiments of the present disclosure. In particular, Figure 11 is a diagram showing an example of a plurality of spots 1142-1148 with coded DNA 1002-1008 (after writing codes) attached to a wafer (or other substrate) shown as a flat surface 1101, and a process for removing, storing and reading the data written at each spot (Spot1-SpotN). In particular, referring to Figure 33, a diagram showing an example of a plurality of spots with coded DNA (after writing codes) attached to a wafer and a process for removing, storing and reading the data written at each spot is shown, in accordance with embodiments of the present invention. After the desired codes are written to the DNA memory strings (or strands or nackets) 1002-1008 for each of the spots 1142-1148 with having the coded DNA memory strings 1002-1008 attached, can be unloaded and the coded DNA memory strings detached or removed from their respective spots as discussed herein and in the aforementioned patent applications. In some embodiments, there may be a plurality of coded DNA memory strings attached to a given spot (as discussed herein above). The detached coded DNA memory strings are then fluidically transported (shown by arrow 1110) along an output channel to a collection bin or container 1112 which holds the coded DNA strings from all the spots in a given wafer array outside of (or separate from) the wafer. When it is desired to read the stored data, the coded DNA memory strings in the collection bin 1112 may be read by any known off-the-shelf DNA reader/sequencer 1114 (such as DNA sequencers made by Illumina or Oxford Nanopore or others) having an accuracy sufficient to meet the needs of the desired application and to determine the DNA sequences written on each of DNA memory strings.

[0182] The DNA reader/sequencer 1114 may provide the code data values from the memory strings to a computer-based system 1126 which performs a decoding and mixture confirmation logic 1127 (which may be implemented by the flow diagram 3400 discussed hereinafter with Fig. 34), which analyzes and decodes the data from the DNA sequencer and confirms it is authentic based on the cassette mixture/formulation for a given lot number, per the binary code to cassette tables (Figs. 29A-29C) discussed herein. The computer system may be such as that described herein in Fig. 30B or similar. The computer system 1126 may communicate with a DNA data server 1124 (which may be the same as or similar to the DNA data server 1915 of Fig. 30A), which

may have the binary code to cassette tables by lot number stored for use by the decoding and mixture confirmation logic 1127. In some embodiments, the DNA Sequencer may save the code data directly to the DNA data server 1124, where it may be retrieved by the decoding and mixture confirmation logic 1127. In some embodiments, the computer system 1126 may communicate with a display 1125, which may display or report data results to the user from reading the DNA encoded data memory strings 1100.

[0183] In some embodiments, the data may be written to the DNA string using a format of address/data, similar to that shown in Fig. 35B, where the address or number of the spot being written to is coded, followed by the data associated with that address (or spot number). Other formatting may be used if desired. Each spot is populated with a plurality of DNA starter (or acceptor) strings (as discussed herein) and they may all be written simultaneously. The number of DNA strings or strands per spot will depend on the liquid spot size and may range from thousands to billions of DNA strings or strands per spot, and other quantities of DNA strings may be used if desired. Also, in some embodiments, for applications where the spot address is not important, e.g., if the coded DNA is left on the array the spot address need not be used as part of the code.

[0184] Referring to Figure 34, a flow diagram 3400 is shown for implementing the decoding and mixture confirmation logic 1127 (Fig 33) which decodes and confirms polymer memory string data, in accordance with embodiments of the present disclosure. The logic 3400 begins at block 3402 by retrieving the Lot# for the wafer array or chip, which may be printed on the wafer or otherwise associated with the wafer 10 (Fig. 23). It also retrieves the DNA bases data from the DNA sequencer's read of all the memory strings on the wafer, e.g, from the DNA Data Server 1124. The logic also retrieves the Binary Code to Cassette Table from the DNA Data Server 1124. Next, block 3404 of the logic separates the memory strings or nackets by address or ID and identifies the cassettes along each string using topo spacing (discussed hereabove). Then, blocks 3406, 3408, 3410, 3412, 3414, 3416, 3418, 3420 of the logic identifies the cassettes in a given string and analyzes the cassettes for one assigned to bit codes per the binary code to cassette table, such as that shown in Figs. 29A, 29B, 29C. If there is a match, a counter for that bit code is incremented as shown by blocks 3408, 3412, 3416, 3418. The process repeats via blocks 3422, 3424 until all cassettes for a given memory string or nacket are reviewed. Then, block 3428 of the logic arranges the 2-bit codes based on writing direction determined from reading the Binary Code to Cassette Table or from the end cap flag for that memory string or nacket. Next, block 3430 of

the logic performs determines if all the memory strings with the current address or Nacket ID have been decided. If No, block 3432 gets the next string/Nacket and the process repeats with block 3406 for all the memory strings with the same address or ID until all complete. Then, when complete, block 3436 of the logic determines if the counter number for each 2-bit code matches the expected distribution (or proportion) of cassettes for that code based on the lot number for a given memory string or nacket address or ID. If it matches, block 3440 of the logic sets a confirmation flag to Pass which confirms the data is authentic for a given memory string or nacket address or ID. If it does not match, block 3438 of the logic sets a confirmation flag to Fail to flag it as a fail status and thus the data is erroneous or counterfeit. Next, block 3442 of the logic checks if all the memory strings/nacket addresses or IDs have been decoded and verified. If not, block 3444 of the logic gets the next string or nacket address/ID and the process repeats with block 3406 for the next address until all have been decoded and verified and the result of block 3442 is Yes. Then the logic exits.

[0185] Figures 35A and 35B are diagrams showing examples of cassettes making up address, data, and error checking for written DNA/polymer memory strings, in accordance with embodiments of the present disclosure. In particular, referring to Figs. 35A and 35B, the format of how data written to the memory string may vary based on various factors and design criteria. In particular, the "memory string" (or memory strand or DNA or polymer or nacket or chain) 1802 may be shown as a line on which are a series of ovals 1804, indicative of individual cassettes written (or added) onto the memory string in a given memory cell, where a cassette is indicative or represents one or more binary (or other radix) bits, depending on the desired encoding scheme, as discussed herein. In some embodiments, the cassette (or bits) 1802 may be written one after the other to build a "storage word". A first example data format shows three components to the storage word, an address section, a data section, and an error checking section. The address section may be a label or pointer used by the memory system to locate the desired data. Unlike traditional semiconductor memory storage where hardware address lines on a computer memory bus would address a unique memory location on the physical memory chip, the memory strings of the present disclosure may have the address (or label) be part of the data stored and indicative of where the data desired to be retrieved is located. In the examples shown in Figs. 35A and 35B, the address for the data written to each spot on a substrate or wafer is located proximate to or contiguous with the data, as well as error checking data, such as parity, checksum, error correction code (ECC), cyclic redundancy

check (CRC), or any other form of error checking and/or security information, including encryption information. In the storage word, each of the components Address, Data, Error Checking, are located after each other in the memory string. As each of the components have a known length (number of bits), e.g., address = 32 bits, data = 16 bits, error check = 8 bits, each storage word and its components can be determined by counting the number of bits. Also, as discussed herein and in the aforementioned commonly owned issued patent and patent applications, a given bit may be represented by one or more NDA bases or oligomers or the like (e.g., a cassette). When a plurality of bases are used to represent one or more bits (e.g., 0,1 or 00,01,10,11, or the like, for a binary system, or G, C, A, T, for a base 4 system), they may be referred to as a "cassette", as discussed herein. Thus, as used herein, the term bit and cassette may be used interchangeably. In some embodiments, there may be a plurality of digital words (address, data, error checking) stored on a given DNA memory string, depending on how long the DNA string can be written.

[0186] Referring to Fig. 35A, an example data format shows the same three components, address section, data section, and error checking section. However, in some embodiments, in between each of the sections there may be a "special bit(s) or sequence" sections S1, S2, S3, as shown by the string 1812. These special bits S1, S2, S3 may be a predetermined series of bits or code that indicate what section is coming next, e.g., 1001001001 may indicate the address is coming next, whereas 10101010 may indicate the data is coming next, and 1100110011 may indicate the error checking section is next. In some embodiments, the special bits may be a different molecular bit or bit structure attached to the string, such as dumbbell, flower, or other "large" molecular structure that is easily definable when the DNA memory string is read offline, outside of the nano-writing chip described herein. Instead of it being large, it may have other molecular properties that provide a unique change to the polymer construction for the bit values. Any other data formatting approaches may be used if desired for the memory strings.

[0187] Figure 36A is a diagram showing a method for creating unique cryptographic DNA fingerprints, in accordance with embodiments of the present disclosure. In particular, individual variability and uniqueness of the originally synthesized (or written) DNA 3602 can be further enhanced by taking the full set of molecules synthesized 3602 and amplifying a collection (or sample or group) of them in separate PCR reactions. Each PCR amplification reaction introduces inherent bias. Also, with each PCR reaction, a different subset of molecules 3603, 3609, 3615, in

the original mix 3602 will be preferentially amplified, shown as PCR Reactions 1-3, 3604, 3610, 3616, respectively, resulting in distinct molecular fingerprints 3606, 3612, 3618, respectively, as shown in Figure 36A. These molecular fingerprints 3606, 3612, 3618 can then be used to create customized molecular codes to incorporate into individual objects or sets of objects or for other secure data purposes.

[0188] Figure 36B is a diagram showing three layers of data derived from a common DNA sequence, in accordance with embodiments of the present disclosure. In particular, the diagram shows how, in some embodiments, each read of a molecular code 3650 may generate 3 layers of data: a binary layer 3652, a production log fingerprint layer 3654 and an object fingerprint layer 3656. The binary layer 3652 is unchanging and may be permanently linked with a Blockchain or NFT hash or any other secure traceable database. The production lot fingerprint layer 3654 is determined by measurements of the % abundance (or proportions) of the different DNA cassette variants used in writing the bits. The original fingerprint 3650 may be stored on the blockchain. The object fingerprint layer 3656 may be viewed as a list of random numbers from each read, where the decoding sequence has unique values. A certain number of numbers during verification must match those originally found to provide authentication. In some embodiments, all three layers 3652, 3654, 3656 are in the same DNA sequence and are inseparable. In some embodiments, the top layer 3652 enables the sequence to be tied to a blockchain, where the blockchain contains encrypted information to validate the other two layers. If using a public blockchain, the layers will survive even if the maker of code goes out of business or ceases to exist, as the DNA will always be readable long into the future.

[0189] In some embodiments, in addition to performing the PCR fingerprints shown in Fig. 36A, additional steps may be performed to provide additional protection against unauthorized copying of the code. For example, a small sample or “seed” of original DNA may be mixed into the end batch. In particular, in some embodiments, the originally synthesized (or written) DNA (or a portion thereof) may be collected in a collection vessel or vial, which are all unique, and a sample extracted and PCR amplified to create a unique fingerprint as discussed herein with Fig. 36A. Also, in parallel, a small sample or “seed” of the unique batch is not amplified and added to the resulting output mixture. The resulting output mixture will have the unique fingerprint but will also have the unique seed sequences, which should not have any duplication (unlike the PCR amplified

sample). Such an approach would reveal if a third party tried to duplicate the process by amplifying the entire sample (including the seed), which would fail the validation check.

[0190] For example, in some embodiments, the output mixture, which may be further incorporated into an object, comprises a sample of the molecules as directly written. This sample of molecules as directly written may further be amplified, e.g., by PCR, wherein the amplification process may introduce a bias artifact into the relative proportions of the original molecules, yielding a unique mixture and associated fingerprint. In some embodiments, an output mixture comprising amplified sequences may further comprise a seed of a original DNA mixture, e.g., wherein the sequences of the original DNA mixture are only present as single copies. In such cases, the output mixture is resistant to amplification attacks, wherein an informed analysis of an output mixture (e.g., sampled from sequences incorporated and subsequently extracted from a suspected counterfeit object) will detect and provide evidence of if an unauthorized third party sampled and amplified the output mixture, e.g., to incorporate into non-authentic or counterfeit objects; such unauthorized interaction with the output mixture will be evidenced in validation analysis, e.g., wherein the counterfeit output mixture comprises multiple copies of the original seed molecules. In some embodiments, the sample of molecules as directly written ("sample molecules"), optionally amplified, and the seed molecules may be of different lengths or different numbers of nucleotides. In some embodiments, the sample molecules, optionally amplified, and the seed molecules may comprise different end cap moieties, e.g., such that primers/probes can index which molecules to read. In some embodiments, the sample molecules and seed molecules are produced in the same, different, or multiple production lots or reactions. In some embodiments, the sample molecules and seed molecules comprise the same or different number or composition of cassettes. In some embodiments, the sample molecules and seed molecules comprise the same or different chemical moieties at the ends of the molecules, and/or the same or different chemical moieties incorporated within the nucleotide backbone, and/or the same or different chemical moieties decorating the nucleotides within the cassettes. In some embodiments, the sample molecules and seed molecules are incorporated together into beads, e.g., silica beads. In some embodiments, the sample molecules and seed molecules are distinctly incorporated into beads, e.g., silica beads, such as in different populations of beads, or in the same population of beads but in different sub-aspects of the beads, e.g., wherein the sample molecules are within the silica beads and the seed molecules are adsorbed to the outside surface of the silica beads, or vice-versa.

[0191] Figure 37 is a diagram showing a method for encoding/decoding system for encoding and decoding a digital file to and from DNA, in accordance with embodiments of the present disclosure.

[0192] Figures 38A, 38B, 38C, 38D, 38E, 38F are diagrams showing a method for the system of Fig. 37 for encoding a digital file into DNA for writing, in accordance with embodiments of the present disclosure. In particular, data from a raw digital file is broken into blocks (B) after data is prepended with the file length and padded to next block size. Then each block (B) is broken into "Nackets" or Nucleic Acid Packets, as each DNA memory string or nacket can only hold a certain amount number of bases, which corresponds to a certain number of bytes of data. For example, a memory string or nacket may hold about 650-2000 DNA bases, and a cassette may be about 20-22 bases long, which would mean a memory string or nacket may range from about 32-100 cassettes long, other values may be used based on the chemistry. Thus, 32 cassettes and 2 bits per cassette, one memory string or nacket may represent only 64 bits or 8 bytes (assuming 8 bits/byte). Each block (B) may be prepended with a block level CRC (cyclic redundancy check), e.g., CRC 32 on the block, and then broken into Data Payloads (W). Next, Parity Nacket Payloads (Z) are calculated, e.g., with ZEFC standard library, which are added to the Nackets, other CRCs may be used if desired. The result is an output of Nacket Payloads (Y) to the next stage, where $Y = W + Z$. Finally, each Nacket is given an Nacket ID (NID) where the total nackets are $N = B(W+Z)$. Also, nacket CRC is calculated based on the Nacket ID and Nacket Payload combined. Thus the final binary digital code being written for a given memory string or nacket will be [Nacket ID][CRC][Nacket Payload], as shown in Fig. 38E. Next, the system may use the approach discussed herein to convert the desired Binary Code to a memory string or nacket to be written to the surface of a wafer array or chip, using an inkjet DNA writing system discussed herein or other DNA synthesis system.

[0193] Figures 39A, 39B, 39C, 39D, 39E, 39F, 39G are diagrams showing a method for the system of Fig. 37 for decoding written DNA back into the original digital file, in accordance with embodiments of the present disclosure. In particular, the encoding process is reversed and data is extracted and determined if the nackets are valid and validated using the CRC. Output nackets can be put into two buckets or classified with a quality score. Low quality nackets may have multiple payloads. Next the nackets are assembled into a block (B) use error correction to determine the original block, e.g., ZEFC, SHA, or MD5 reconstruction. Then, verify the block with CRC and

repeat the process across all nackets to obtain all the blocks in the read. Then, the blocks are re-assembled and the original raw data file is obtained.

[0194] Figures 40A, 40B, 40C are data graphs showing results data using the encode/decode system of Fig. 37, in accordance with embodiments of the present disclosure. In particular, Fig. 40A shows a Nacket classification bar graph (or histogram) 4000 showing number of Nacket reads (log scale) on the Y axis vs Nacket address (or Nacket ID) on the X axis. This data shows a large number of full length and correct CRC and consensus Nackets were found for a 200 byte test. Fig. 40B shows pie charts 4050, 4052 for 3.5Kbyte test, there the left pie chart 4050 showing a breakdown of all reads to valid reads (9.94%), and the pie chart 4052 on the right shows a breakdown of full Nacket ID, Symbol Free Reads, Full Length Nackets, Full Length and Correct CRC Nackets, and Full Length and Correct CRC and Consensus Nackets. Fig. 40C shows a bar graph (or histogram) 4060 showing number of Nacket reads (log scale) on the Y axis vs Nacket address (or ID) on the X axis (similar to Fig. 40A) for the 3.5Kbyte test.

[0195] Figure 50 shows an alternative embodiment for writing randomly selected mixtures of cassettes using a computer (or CPU) generated randomness instead of a physical mixture of cassettes. In particular, Figure 50 is a diagram showing print head banks 5010, 5012, 5014, 5016 for a laser jet DNA printer having separate topo cassettes nozzles 5010A, 5012A, 5014A, 5016A corresponding to each head bank, in accordance with embodiments of the present disclosure. The head banks 5010, 5012, 5014, 5016 are controlled by a print head control logic (or controller) 5020 which selects the appropriate head (within the print head) to write to a spot 14 on the chip or wafer. In this case, there are 4 cassettes associated with each 2-bit code (C1-C4 for "00", C5-C8 for "01", C9-C12 for "10", C13-C16 for "11"). As discussed herein below, the controller 5020 randomly selects which cassette (among the assigned cassettes) to write using a random selection process performed by the control logic, e.g., QRNG (quantum random number generator), or any other desired random number generator that provides a sufficiently random output from a set of numbers. The logic also keeps track of each C# selected and printed during the writing process. When the writing process for the chip is complete, the logic stores the total number of each writes for each C# associated with each 2-bit pair, and calculates the percentage usage of each C# within each 2-bit pair and stores the Code to Cassette table, which is then used during the authentication process, similar to that performed using the physical mixture approach.

[0196] In that case, each spot will have one-dimensional randomness of cassettes along the length of the memory string, instead of two-dimensional randomness for the physical mixture approach discussed with Fig. 28. Accordingly, if the number of cassettes along a memory string or nacket is not sufficient to provide authentication, authentication may be performed across a plurality of spots for a chip 5100, such as across one or more rows, as shown in Figure 51A. In that case, the logic calculates percentage usage of each C# within each 2-bit pair for a given row (or group of rows) and stores the result in a row-based Code to Cassette table 5102, each row having computer-generated random proportions of cassettes (Cs) associated with each two-bit code.

[0197] In some embodiments, the logic may calculate percentage usage of each C# within each 2-bit pair for the entire chip or array as shown in Figure 51B. In that case, the logic calculates percentage usage of each C# within each 2-bit pair for entire chip 5110 and stores the result in a chip-based Code to Cassette table 5112, the entire chip having computer-generated random proportions of cassettes (Cs) associated with each two-bit code for the entire chip, and each chip or lot number may be a different set of proportions.

[0198] Referring to Figure 52, in some embodiments, the print head bank 5200 may have all the cassettes for the entire chip with separate cassette nozzles, e.g., C1-C16, each individually addressable by the controller 5020. In that case, the control logic 5020 determines the desired cassette C1-C16 to write based on the cassette assignment for each 2-bit code and selects that cassette for writing and performs the write. The logic may be similar to that described herein above for Fig. 50 except that, in some embodiments, there would only need a single control line instead of multiple control lines and multiple print head banks.

[0199] Referring to Figure 53A, a flow diagram is shown for writing (printing) and unloading coded polymer memory strings in an inkjet writing system using computer-based randomness for cassette writing selection, in accordance with embodiments of the present disclosure. In particular, this logic is similar to the logic of Fig. 31A having blocks 3102 to 3132, except that instead of retrieving 4 ink cartridges each with a different mixture, it retrieves the heads having the assigned group of individual cassettes, shown as block 5302 (instead of block 3106). Also, block 5304 is provided for writing/printing the 2-bit code which references the writing process in Fig. 53B (instead of block 3114 which referenced the writing process in Fig. 31B).

[0200] Referring to Figure 53B, a flow diagram is shown for writing (printing) 2-bit code to DNA/polymer memory string in an inkjet writing system using computer-based randomness for

cassette writing selection, in accordance with embodiments of the present disclosure. In particular, this logic is similar to the logic of Fig. 31B, except that instead of printing the bits using the pre-set mixture cartridges, corresponding block 5354, 5358, 5362, 5366 of the logic obtains the cartridge with the appropriate Cs for the 2-bit code to be written, then randomly selects the C# from among the Cs assigned to that 2-bit code. Next, corresponding block 5354, 5358, 5362, 5366 prints the corresponding 2-bit code with the randomly selected C# at the desired spot/location on the array or chip. Then the corresponding block 5354, 5358, 5362, 5366 increments the corresponding C# counter for the chip and/or row being written. The process continues until the bit writing is complete for the spot or group of spots being written, as determined by block 5368. When complete, the result of block 5368 is Yes and block 5370 saves the C# counters in the Code-to-Cassette table. Next, block 5372 determines whether there are any droplet errors were detected by the droplet viewer (or sensor) 1911 (Fig. 30A), which may be part of the print head and array stage controller and inspection logic 1908 (Fig. 30A). If Yes, errors were detected and block 5374 saves the error location(s) and bit number for future reading, and the logic exits. If the result of block 5372 is No, then no droplet errors were found for that write cycle and the logic exits. There may be a pre-set Nacket ID to Row conversion that is used for this process. In particular, if each row has a predetermined number of spots, and a pre-set number of redundant spots for error protection, the system can have a pre-determined number of rows (or corresponding Nackets) that will provide sufficient authentication of the cassette C# proportion validation. Also, the logic 5350 of Fig. 53B also stores and updates the C# counters in the Code-to-Cassette Table after each spot is written for future use during authentication.

[0201] Referring to Figure 54, a flow diagram 5400 is shown for decoding and confirming polymer memory string data when using computer-based randomness for cassette writing selection, in accordance with embodiments of the present disclosure. In particular, the logic 5400 is similar to the logic 3400 of Fig. 34, having blocks 3402 to 3440, except that instead of checking the authentication after each Nacket ID is decoded, the logic waits until all the memory strings/Nacket IDs (or at least until the number of row or Nacket IDs used for validation) have been decoded, and then checks the C# counters for the proportions to determine a pass/fail for the ID, Rows, or chip, as shown by block 3442 is performed after block 3430.

[0202] It should be understood that the surface of the substrate being written may be flat (unpatterned) or patterned.

[0203] In some embodiments, the present disclosure may be used with NFTs, Tokens, Contract addresses, pKI components, Digital certs, Private database identifiers, ERP database identifiers, UDIs - for new device, Global trade numbers, GTIN, UPC codes, QR codes, EAN, ISBNs, Library of congress numbers, FNSKU, ITF-14, Contract IDs, for example DOD, Dod CIC credentials, Patient identifiers, EMR records, such as Epic patient IDs, Contractor license numbers, Professional license numbers, Notary identification numbers, Permit numbers for construction, Inspector IDs numbers for construction or QC. The present disclosure may also be used with physical currency (paper, metal, and the like) as well as digital currency, including cryptocurrency such as Payment Cryptocurrencies, Coins, Stablecoins, and Central Bank Digital Currencies, and including Bitcoin, Ethereum, Tether, XRP, Binance Coin, USD Coin, Cardano, Solana, Dogecoin, Tron, Polygon, and the like, including but not limited to other cryptocurrencies now known or later discovered or developed, that may use their own independent blockchain. In addition, the system and method of the present disclosure may authenticate an object by being able to retain, lookup, or validate the production and/or molecular fingerprints, which may be done in a common database or in a separate authentication database that may be hashed or use other/additional encryption or be clear text. In addition, as discussed herein, in some embodiments, the data encoded by the present disclosure may be an NFT and the authentication data and/or encoding data may be on a blockchain.

[0204] In some embodiments, the disclosure provides a method of object authentication according to Method 1 (Method 1A), wherein the nackets are synthesized using an inkjet printing head (e.g. a piezoelectric print head), by sequential addition of cassettes to DNA receptor strands, wherein each cassette comprises multiple nucleotides, wherein in each sequential addition step the cassettes comprise a heterologous population of cassettes of at least two different sequences encoding the same data in a machine-readable code (e.g., binary or ternary code), and wherein the cassettes are dispensed by an inkjet writing print head on at least one writing spot on a wafer array, the head or nozzle writing the same code to a plurality of polymer memory strands dispensed on the at least one spot, e.g. comprising the following steps:

- a) loading the desired spot to be written with a starter polymer or DNA attached at one end to the desired spot;
- b) washing the surface of the spot;

- c) positioning an inkjet nozzle having a heterologous population of cassettes wherein the population comprises cassettes having at least two different sequences, but all encoding the same information in one or more bits (e.g., 1 or 0, or 00, 01, 10, 11, etc. in binary code) over the desired spot to be written corresponding to the unique code;
- d) causing the inkjet nozzle to release a droplet comprising the heterologous population of cassettes onto the spot, thereby writing a bit or portion of the unique code to the DNA or polymer memory strings (or strands) associated with the spot; and
- e) washing the surface of the spot;

optionally further comprising steps f) – i):

- f) causing the inkjet nozzle to release a droplet of deblock/adaptor reagent onto the spot;
- g) washing the surface of the spot; and
- h) repeating steps (c) through (g) until the unique code has been written in the memory string at the spot.
- i) removing the memory strings from the spot and flowing the memory strings from the spot into a collection or storage container for later incorporation into or onto an object

[0205] For example, in the preceding method, the cassettes may be added by topoisomerase mediated ligation, for example by:

- (i) reacting double-stranded acceptor DNA strands with topoisomerases charged with double-stranded DNA cassettes from the heterologous population of cassettes covalently bound to the topoisomerases,

wherein a strand of the acceptor DNA has a 5' overhang,

wherein each cassette comprises an informational sequence, a topoisomerase recognition sequence, and 5' overhangs on both strands,

wherein the 5' overhang of the strand of the oligomer that does not bear the topoisomerase ("bottom strand") is complementary to the 5' overhang of the acceptor DNA but is not complementary to the 5' overhang of the strand bearing the topoisomerase ("top strand") of the cassette,

wherein the 5' end of the strand bearing the topoisomerase ("top strand") of the cassette and 5' end of the acceptor DNA are not protected, e.g., not phosphorylated (i.e., 5'-OH), and

wherein the topoisomerase charged with a double-stranded DNA cassette is delivered to the location of the acceptor strand by a piezo-electric inkjet nozzle;

(ii) reacting the acceptor DNA thus extended in step (i) with a topoisomerase charged with a further double-stranded DNA cassette,

wherein the further cassette comprises an informational sequence that is the same as or is different from any informational sequence in the cassette of step (i), a topoisomerase recognition sequence, and 5' overhangs on both strands,

wherein the 5' overhang of the strand of the further cassette not bearing the topoisomerase ("bottom strand") is complementary to the 5' overhang of the extended acceptor DNA but is not complementary to the 5' overhang of the strand of the further cassette bearing the topoisomerase ("top strand"), and

wherein the 5' end of the strand bearing the topoisomerase ("top strand") of the further cassette is not protected, e.g., not phosphorylated (i.e., 5'-OH); and

(iii) repeating steps (i) and (ii) until the desired nucleotide sequence is obtained; wherein there is optionally a washing step after step (i) and/or after step (ii).

[0206] For example, in some embodiments, the present disclosure provides a method for writing a desired binary code using a DNA or polymer strand or memory string, the desired binary code having a plurality of 2-bit binary codes, comprising: providing a plurality of unique DNA Cassettes for writing four different 2-bit binary codes, a predetermined unique set of the plurality of DNA cassettes being associated with each of the four 2-bit binary codes, each DNA cassette having a same DNA cassette length defined by a predetermined number of positions, each position comprising one of four DNA or polymer bases; providing four inkjet cartridges, each inkjet cartridge associated with a different one of the 2-bit binary codes, and each cartridge having a fluid with a different predetermined DNA cassette mixture of the set of DNA cassettes associated with a given 2-bit binary code; wherein the predetermined DNA cassette mixture being associated with a current lot number or date code; obtaining a first 2-bit binary code from a desired binary code to be written on a surface of a substrate; writing the first 2-bit binary code by applying a droplet of

fluid from the inkjet cartridge associated with the first 2-bit binary code at a memory spot writing location on the surface of the substrate, the droplet comprising the DNA cassettes associated with the first 2-bit binary code; wherein the DNA cassettes in the droplet attach to existing DNA cassettes on the surface in a random arrangement on the surface, the random arrangement being based at least on DNA cassette attachment kinetics and the DNA cassettes in the droplet; repeating the obtaining and writing steps for successive 2-bit binary codes until the desired binary code is written for a given memory spot on the surface of the substrate, wherein each writing step produces a random arrangement of the DNA cassettes associated with the current 2-bit binary code being attached to existing DNA cassettes on the surface, creating a plurality of memory strings at a given memory spot; wherein the total distribution of all the DNA cassettes associated with a given 2-bit binary code in all the memory strings is substantially equal to the predetermined DNA cassette mixture within a predetermined tolerance; and wherein the DNA cassettes associated with a given 2-bit binary code are randomly distributed along a given memory string.

[0207] In addition, in some embodiments, the unique set of the plurality of DNA cassettes associated with each of the four 2-bit binary codes changes for each lot or time code. Also, in some embodiments, the predetermined unique set of the plurality of DNA cassettes being associated with each of the four 2-bit binary codes comprises a unique set of four. Also, in some embodiments, the plurality of unique DNA Cassettes for writing four different 2-bit binary codes comprises 16 unique DNA Cassettes. Also, in some embodiments, the plurality of unique DNA Cassettes for writing four different 2-bit binary codes comprises an integer greater than 2. Also, in some embodiments, the unique set of the plurality of DNA cassettes is associated with each of the four 2-bit binary codes.

[0208] Also, in some embodiments, the number of positions for the DNA cassette length comprises an integer greater than 3. Also, in some embodiments, each position comprises one of four DNA bases plus additional polymer objects, wherein each position comprises one of at least five unique polymer objects. Also, in some embodiments, a first existing DNA cassette comprises a starter cassette or target sequence which is not part of the desired binary code to be written. Also, in some embodiments, the DNA cassettes comprises topo-cassettes having a topoisomerase portion and a cassette binary code portion. Also, in some embodiments, the 2-bit binary code comprises an n-bit binary code. Also, in some embodiments, the predetermined DNA cassette mixture for each of the

2-bit binary codes is derived from the lot number or date code. Also, in some embodiments, the 2-bit binary codes may be an n-bit binary code.

[0209] In another aspect, the disclosure provides a method of synthesizing DNA, e.g., any of DNA 2, et seq., wherein the DNA comprises transitions between non-identical nucleotides corresponding to a series of bits in a machine-readable code, e.g., a ternary code, comprising stepwise addition of nucleotides (dNTPs) into a kinetically controlled reaction mixture comprising one or more transferase, e.g., terminal deoxynucleotidyl transferase (TdT) and one or more dNTP degrading enzymes, e.g., apyrase, wherein each stepwise addition uses a different nucleotide. In this method, at each addition step an indeterminate plurality of nucleotides (e.g., ca. 5-15, with optimal balance of the TdT and apyrase) are added to each strand, before the dNTPs are consumed by the apyrase, then a different dNTP is added, so the strands created have varying lengths, and the data is encoded in the transitions between the nonidentical nucleotides, which is the same for each strand, providing a population of heterologous nucleic acid data packets (“nackets”), wherein each nacket contains a plurality of DNA molecules encoding the same data (here, at the junctions between non-identical nucleotides), wherein the sequences of the DNA molecules are heterogeneous (here, because the lengths of the runs of identical nucleotides is variable). Using the four natural dNTPs, there are three possible transitions for each nucleotide, e.g. AT/AC/AG, TA/TC/TG, CA/CG/CT, and GC/GA/GT. This possibility allows for further synonymous heterologous sequences, as using a ternary code with 0, 1, and 2, each of 0, 1, and 2 could be represented by any of four different transitions (see, e.g., one possible set of permutations at Fig. 41).

[0210] The disclosure further provides a method of decoding the population of DNA molecules; for example, the sequencing of the population of DNA molecules, the compressing of the DNA molecule sequences by filtering out the sequences of identical nucleotides to provide a compressed representative sequence, and using the schema used during data encoding to decode the compressed representative sequence back into the original data string. Alternatively, or additionally, the sequences of the population of DNA molecules may be further analyzed using statistical inference methods and/or models, such as those disclosed in Lee, H.H., *et al.*, “*Terminator-free template-independent enzymatic DNA synthesis for digital information storage.*” Nat. Commun. (2019)10:2383, the contents of which are incorporated herein by reference.

[0211] For example, the disclosure provides a method (Method 2) for writing a desired code, e.g., a ternary code, using a DNA strand, comprising:

- i. providing a reaction mixture comprising one or more transferase enzyme, e.g., terminal deoxynucleotidyl transferase (TdT) and one or more dNTP degrading enzyme, e.g., apyrase;
- ii. adding to the reaction mixture a deoxyribonucleotide triphosphate (dNTP);
- iii. waiting until the dNTP of step (ii) is added to the DNA strand or degraded;
- iv. repeating steps (ii) and (iii) until the desired bit sequence is reached, wherein non-identical dNTP species are used in any two consecutive additions

thereby providing a population of DNA molecules encoding the desired data string.

For example, in particular embodiments the disclosure provides:

2.1. Method 2, further comprising the steps of

- v. optionally, storing the reaction mixture for further addition(s), purification, or processing;
- vi. purifying the synthesized DNA or polymer strand or memory string comprising the data string; and
- vii. optionally, storing purified DNA or polymer strand or memory string for later use, analysis, addition(s), purification, or processing.

2.2. Any foregoing Method, wherein the reaction mixture comprises terminal deoxynucleotidyl transferase (TdT).

2.3. Any foregoing Method, wherein the reaction mixture further comprises apyrase.

2.4. Any foregoing Method, wherein the reaction mixture is aqueous, e.g., a buffer.

2.5. Any foregoing Method, wherein the reaction mixture further comprises further additives, e.g., ions, e.g., cations, e.g., divalent cations, e.g., cobalt.

2.6. Any foregoing Method, wherein the reaction mixture comprises a mixture of TdT and apyrase, e.g., in a stoichiometric ratio such that kinetically-controlled stepwise addition of dNTPs is achieved.

2.7. Any foregoing Method, wherein the dNTPs comprise adenosine triphosphate (ATP), guanosine triphosphate (GTP), cytidine triphosphate (CTP), thymidine triphosphate (TTP); optionally, uridine triphosphate (UTP).

- 2.8. Any foregoing Method, wherein the 3-bit ternary code comprises an n-bit ternary code.
- 2.9. Any foregoing Method, wherein the synthesized DNA or polymer strand or memory string, or the population of DNA molecules synthesized, comprises any of DNA 2, et seq.
- 2.10. Any previous method, for use in combination with any of the methods of Methods 1, et seq., Methods 3, et seq., Methods 4, et seq., Methods 5, et seq., Methods 6, et seq., and/or Methods 7, et seq.

[0212] The disclosure thus provides a method of object authentication (Method 3), comprising:

- i. synthesizing DNA sequences comprising nucleic acid data packets (“nackets”), wherein each nacket contains a plurality of DNA molecules encoding the same data, wherein the sequences of the DNA molecules are synthesized using one or more transferase enzyme, e.g., terminal deoxynucleotidyl transferase (TdT);
- ii. incorporating said DNA sequences into or onto an object;
- iii. extracting said DNA sequences from the object; and
- iv. analyzing the extracted DNA sequences;
- v. optionally, comparing the analyzed DNA sequences to a database of DNA sequences;
- vi. optionally, confirming object authenticity.

[0213] For example, in particular embodiments the disclosure provides:

- 3.1. Method 3, wherein the DNA sequences encodes data that functions as an identification code for the object.
- 3.2. Method 3.1, wherein the data that functions as an identification code is randomly generated.
- 3.3. Any previous method, wherein the DNA sequences comprise any of DNA 2, et seq.
- 3.4. Any previous method, wherein the DNA sequences are synthesized by sequential addition of homopolymer extensions, wherein each subsequent homopolymer extension comprises a non-identical nucleotide from the adjacent homopolymer extension(s).

- 3.5. Method 3.4, wherein the homopolymer extensions are synthesized using a transferase enzyme, e.g., terminal deoxynucleotidyl transferase (TdT).
- 3.6. Method 3.4, wherein the homopolymer extensions are synthesized using TdT.
- 3.7. Any previous method, wherein the DNA sequences are incorporated into an object by direct surface conjugation of the DNA sequences onto the object.
- 3.8. Any previous method, wherein the DNA sequences are incorporated into a constituent part or material of an object used in production of said object, optionally into textiles, fabrics, leather, biomaterial products, polymers, plastics, wood, metals, inks, paints, solutions, suspensions, and raw materials.
- 3.9. Any previous method, wherein the DNA sequences are encapsulated into a micro-container, optionally a microsphere, optionally a silica microsphere, prior to incorporation into the object.
- 3.10. Any previous method, wherein the DNA sequences are encapsulated into a molecular assembly, such as a lipid nanoparticle, protein complex or aggregate, or crystal lattice.
- 3.11. Any previous method, wherein the DNA sequences are inserted into a cell or cells, optionally inserted into a larger DNA construct and/or genome, optionally inserted into yeast, bacteria, fungi, plant, or animal cells, optionally wherein the cells are used in the production of foods, drinks, biologics, or materials, e.g., cheese, beer, wine, vegan leather, pharmaceuticals.
- 3.12. Any previous method, wherein the incorporated DNA sequences are extracted from the object through physical means, optionally cutting, grinding, scoring, chipping, shredding, or pulverizing one or more pieces of the object.
- 3.13. Any previous method, wherein the incorporated DNA sequences are extracted from the object through chemical means, optionally dissolving or cleaving the DNA sequences and/or one or more pieces of the object.
- 3.14. Any previous method, wherein the extracted DNA sequences are isolated and/or purified by chromatography, e.g., ion exchange chromatography, size exclusion chromatography, normal-phase or reverse-phase high-performance liquid chromatography (HPLC), antibody affinity chromatography, or combinations thereof.

- 3.15. Any previous method, wherein the extracted DNA sequences are isolated and/or purified by immobilization, e.g., solid-phase reversible immobilization (SPRI), immunoprecipitation (or antibody pull-down), or combinations thereof; further optionally in solution, resin, slurry, bead, filter, or combinations thereof.
- 3.16. Any previous method, wherein the extracted DNA sequences are isolated and/or purified by electrophoresis, e.g. polyacrylamide gel electrophoresis, two-dimensional electrophoresis, pulsed field electrophoresis, Southern blotting, or combinations thereof.
- 3.17. Any previous method, wherein the extracted DNA sequences are isolated and/or purified by centrifugation, further optionally by filtration, e.g., spin columns.
- 3.18. Any previous method, wherein the extracted DNA sequences are analyzed using mass spectrometry and/or high-throughput DNA sequencing.
- 3.19. Any previous method, wherein the extracted DNA sequences are compared to a database containing the object identification codes as originally synthesized for said object.
- 3.20. Any previous method, wherein the extracted DNA sequences are compared to results from one or more previous analysis of extracted DNA sequences from the same or similar object.
- 3.21. Any previous method, for use in combination with any of the methods of Methods 1, et seq., Methods 2, et seq., Methods 4, et seq., Methods 5, et seq., Methods 6, et seq., and/or Methods 7, et seq.
- 3.22. Any previous method, wherein the DNA sequences comprise any of DNA 1, et seq., and/or DNA 2, et seq.

[0214] The disclosure thus provides a method for writing an attack-resistant digital code using DNA (Method 4), comprising:

- i. receiving a desired digital code to be written, the desired code being grouped into four two-bit binary codes to be written (e.g., 00, 01, 10, 11);
- ii. providing four predetermined mixtures of a predetermined number of unique DNA cassette strings, each mixture corresponding to a different predetermined two-bit binary code value, each mixture having a predetermined proportion of the unique DNA

- cassettes within the mixture, and the unique DNA cassette strings of each mixture being different from the DNA cassette strings in the other mixtures;
- iii. depositing a droplet of the mixture associated with a given two-bit binary code to be written onto a substrate to add a DNA cassette string to an encoded DNA string being written, the droplet comprising the predetermined mixture of the unique cassettes associated with the given two-bit binary code; and
 - iv. repeating the depositing until the desired code is written onto the encoded DNA string.

[0215] For example, in particular embodiments the disclosure provides:

- 4.1. Method 4, further comprising, after the desired code is written, adding an end cap to the encoded DNA string.
- 4.2. Method 4.1, wherein the end cap contains information about the desired digital code or how to read the code.
- 4.3. Any previous method, wherein the substrate has an acceptor DNA strand having one end attached to the substrate and an opposite end being available to attach to one of the unique DNA cassettes to be added.
- 4.4. Any previous method, wherein the predetermined number of unique DNA cassettes for one of the mixtures is different from at least one other of the mixtures.
- 4.5. Any previous method, wherein the desired digital code is encoded in an NFT with authentication data and stored on a blockchain.
- 4.6. Any previous method, wherein the encoded DNA string is embedded in a physical object to be authenticated.
- 4.7. Any previous method, for use in combination with any of the methods of Methods 1, et seq., Methods 2, et seq., Methods 3, et seq., Methods 5, et seq., Methods 6, et seq., and/or Methods 7, et seq.
- 4.8. Any previous method, wherein the DNA sequences comprise any of DNA 1, et seq., and/or DNA 2, et seq.

[0216] The disclosure thus provides a method for writing an attack-resistant digital code using DNA (Method 5), comprising:

- i. receiving a desired digital code to be written, the desired code being grouped into a plurality of n -bit binary codes to be written, where n is greater than 1;
- ii. providing at least two predetermined mixtures of a predetermined number of unique DNA cassette strings, each mixture corresponding to a different predetermined n -bit binary code value, each mixture having a predetermined proportion of the unique DNA cassettes within the mixture, and the unique DNA cassette strings of each mixture being different from the DNA cassette strings in the other mixtures;
- iii. depositing a droplet of the mixture associated with a given n -bit binary code to be written onto a substrate to add a DNA cassette string to an encoded DNA string being written, the droplet comprising the predetermined mixture of the unique cassettes associated with the given n -bit binary code; and
- iv. repeating the depositing until the desired code is written onto the encoded DNA string.

[0217] For example, in particular embodiments the disclosure provides:

- 5.1. Method 5, further comprising, after the desired code is written, adding an end cap to the encoded DNA string.
- 5.2. Method 5.1, wherein the end cap contains information about the desired digital code or how to read the code.
- 5.3. Any previous method, wherein the substrate has an acceptor DNA strand having one end attached to the substrate and an opposite end being available to attach to one of the unique DNA cassettes to be added.
- 5.4. Any previous method, wherein the predetermined number of unique DNA cassettes for one of the mixtures is different from at least one other of the mixtures.
- 5.5. Any previous method, wherein the desired digital code is encoded in an NFT with authentication data and stored on a blockchain.
- 5.6. Any previous method, wherein the encoded DNA string is embedded in a physical object to be authenticated.
- 5.7. Any previous method, for use in combination with any of the methods of Methods 1, et seq., Methods 2, et seq., Methods 3, et seq., Methods 4, et seq., Methods 6, et seq., and/or Methods 7, et seq.

5.8. Any previous method, wherein the DNA sequences comprise any of DNA 1, et seq., and/or DNA 2, et seq.

[0218] The disclosure thus provides a method for writing an attack-resistant digital code using DNA (Method 6), comprising:

- i. receiving a desired digital code to be written, the desired digital code being grouped into four two-bit binary codes to be written (e.g., 00, 01, 10, 11);
- ii. providing four sets of unique DNA cassette strings, each set comprising a predetermined number of unique DNA cassettes and each set corresponding to a different predetermined two-bit binary code value, such that each set of unique cassettes corresponding to different two-bit binary code and each set of unique cassette strings being different from the other DNA cassette strings;
- iii. randomly selecting one of the unique cassettes corresponding to a given two-bit binary code to be written, as a selected unique cassette;
- iv. depositing a droplet of the selected unique cassette associated with the given two-bit binary code to be written onto a substrate to add the selected unique cassette to an encoded DNA string being written;
- v. repeating the selecting and depositing until the desired code is written onto the encoded DNA string on a given writing spot on the substrate; and
- vi. counting the number of times each unique cassette is used for each two-bit binary code written.

[0219] For example, in particular embodiments the disclosure provides:

- 6.1. Method 6, further comprising, after the desired code is written, adding an end cap to the encoded DNA string.
- 6.2. Method 6.1, wherein the end cap contains information about the desired digital code or how to read the code.
- 6.3. Any previous method, wherein the substrate has an acceptor DNA strand having one end attached to the substrate and an opposite end being available to attach to one of the unique DNA cassettes to be added.

- 6.4. Any previous method, wherein the predetermined number of unique DNA cassettes for one of the mixtures is different from at least one other of the mixtures.
- 6.5. Any previous method, wherein the desired digital code is encoded in an NFT with authentication data and stored on a blockchain.
- 6.6. Any previous method, wherein the encoded DNA string is embedded in a physical object to be authenticated.
- 6.7. Any previous method, for use in combination with any of the methods of Methods 1, et seq., Methods 2, et seq., Methods 3, et seq., Methods 4, et seq., Methods 5, et seq., and/or Methods 7, et seq.
- 6.8. Any previous method, wherein the DNA sequences comprise any of DNA 1, et seq., and/or DNA 2, et seq.

[0220] The disclosure thus provides a method for writing an attack-resistant digital code using DNA (Method 7), comprising:

- i. receiving a desired digital code to be written, the desired digital code being grouped into a plurality of n -bit binary codes to be written, where n is greater than 1;
- ii. providing at least two sets of unique DNA cassette strings, each set comprising a predetermined number of unique DNA cassettes and each set corresponding to a different predetermined n -bit binary code value, such that each set of unique cassettes corresponding to different n -bit binary code and each set of unique cassette strings being different from the other DNA cassette strings;
- iii. randomly selecting one of the unique cassettes corresponding to a given n -bit binary code to be written, as a selected unique cassette;
- iv. depositing a droplet of the selected unique cassette associated with the given n -bit binary code to be written onto a substrate to add the selected unique cassette to an encoded DNA string being written;
- v. repeating the selecting and depositing until the desired code is written onto the encoded DNA string on a given writing spot on the substrate; and
- vi. counting the number of times each unique cassette is used for each n -bit binary code written.

[0221] For example, in particular embodiments the disclosure provides:

- 7.1. Method 7, further comprising, after the desired code is written, adding an end cap to the encoded DNA string.
- 7.2. Method 7.1, wherein the end cap contains information about the desired digital code or how to read the code.
- 7.3. Any previous method, wherein the substrate has an acceptor DNA strand having one end attached to the substrate and an opposite end being available to attach to one of the unique DNA cassettes to be added.
- 7.4. Any previous method, wherein the predetermined number of unique DNA cassettes for one of the mixtures is different from at least one other of the mixtures.
- 7.5. Any previous method, wherein the desired digital code is encoded in an NFT with authentication data and stored on a blockchain.
- 7.6. Any previous method, wherein the encoded DNA string is embedded in a physical object to be authenticated.
- 7.7. Any previous method, for use in combination with any of the methods of Methods 1, et seq., Methods 2, et seq., Methods 3, et seq., Methods 4, et seq., Methods 5, et seq., and/or Methods 6, et seq.
- 7.8. Any previous method, wherein the DNA sequences comprise any of DNA 1, et seq., and/or DNA 2, et seq.

[0222] The system, computers, servers, devices and the like described herein have the necessary electronics, computer processing power, interfaces, memory, hardware, software, firmware, logic/state machines, databases, microprocessors, communication links (wired or wireless), displays or other visual or audio user interfaces, printing devices, and any other input/output interfaces, to provide the functions or achieve the results described herein. Except as otherwise explicitly or implicitly indicated herein, process or method steps described herein may be implemented within software modules (or computer programs) executed on one or more general-purpose computers. Specially designed hardware may alternatively be used to perform certain operations. Accordingly, any of the methods described herein may be performed by hardware, software, or any combination of these approaches. In addition, a computer-readable storage

medium may store thereon instructions that when executed by a machine (such as a computer) result in performance according to any of the embodiments described herein.

[0223] In addition, computers or computer-based devices described herein may include any number of computing devices capable of performing the functions described herein, including but not limited to: tablets, laptop computers, desktop computers, smartphones, mobile communication devices, smart TVs, set-top boxes, e-readers/players, and the like.

[0224] Although the disclosure has been described herein using exemplary techniques, algorithms, or processes for implementing the present disclosure, it should be understood by those skilled in the art that other techniques, algorithms and processes or other combinations and sequences of the techniques, algorithms and processes described herein may be used or performed that achieve the same function(s) and result(s) described herein and which are included within the scope of the present disclosure.

[0225] Any process descriptions, steps, or blocks in process or logic flow diagrams provided herein indicate one potential implementation, do not imply a fixed order, and alternate implementations are included within the scope of the preferred embodiments of the systems and methods described herein in which functions or steps may be deleted or performed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art.

[0226] It should be understood that, unless otherwise explicitly or implicitly indicated herein, any of the features, functions, characteristics, alternatives or modifications described regarding a particular embodiment herein may also be applied, used, or incorporated with any other embodiment described herein. Also, the drawings herein are not drawn to scale, unless indicated otherwise.

[0227] Conditional language, such as, among others, "can," "could," "might," or "may," unless specifically stated otherwise, or otherwise understood within the context as used, is generally intended to convey that certain embodiments could include, but do not require, certain features, elements, or steps. Thus, such conditional language is not generally intended to imply that features, elements, or steps are in any way required for one or more embodiments or that one or more embodiments necessarily include logic for deciding, with or without user input or prompting, whether these features, elements, or steps are included or are to be performed in any particular embodiment.

EXAMPLES

[0228] There are myriad aspects to be considered in the application of DNA for object authentication and object provenance, for example:

- Encoding: The conversion of a machine-readable code, e.g., binary code, e.g., an identification code, e.g., NFT, into DNA, e.g., nackets.
- Accessibility: Incorporating the free DNA strands directly into the object, or optionally encapsulating the DNA, e.g., into silica beads or microspheres.
- Formulation: The method of physically mixing the DNA (free or encapsulated) into the object or material of interest, e.g., a material for subsequent production of the object.
- Application: The method of using the formulated object or material, e.g., applying ink to paper, e.g., applying paint to canvas or drywall, etc.
- Sampling: The method of extracting the DNA (free or encapsulated) from the object or material; optionally, further removing the encapsulated DNA from the encapsulating material, e.g., silica beads or microspheres.
- Reading: The method of DNA analysis, e.g., DNA sequencing; optionally, further comprising one or more amplification steps, e.g., PCR amplification.
- Decoding: The method of, optionally, converting the DNA sequence into the original machine-readable code, i.e., reconstituting the original data file.

EXAMPLE 1: OBJECT AUTHENTICATION USING FOUNTAIN PEN INK

[0229] To exemplify one embodiment of the present disclosure, six commercially-available fountain pen inks of various colors are acquired. Each ink is labeled Ink #1 through Ink #6, and each ink is serially diluted 10-fold four times. Separately, a 32-byte NFT, along with accompanying meta-data and error correcting features, is encoded into DNA strands synthesized using topoisomerase-mediated heterologous DNA cassette data writing, with said DNA strands comprising 51 nackets each. The DNA is added to each of the ink samples (i.e., Ink #1 through Ink #6, across four dilutions each) at a concentration of 0.3 ng/ μ L. As an initial evaluation, the DNA is added to the ink samples, mixed thoroughly, and immediately aliquoted for DNA analysis. The DNA is subsequently isolated and amplified to verify that introduction into the ink is not deleterious in the process of object (i.e., ink) authentication.

[0230] Next, Ink #4 and Ink #5 are selected for further evaluation, since both inks are black inks, though color does not seem to impact the DNA based on the above experiment. NFT-encoding DNA is incorporated into the fountain pen inks as described above, the inks are used in fountain pens to write on commercially-available printer paper, and are subsequently analyzed after 7 days to evaluate the stability of the DNA in both the liquid ink and when written/dried on the paper. The DNA is subsequently isolated and amplified. When sampling directly from the ink solution, an aliquot of the ink solution is diluted and then directly amplified via PCR. When sampling from the ink dried on paper, a wetted cotton swab is lightly brushed over the dried ink, dipped in a small volume of water, and then amplified via PCR. Alternatively, the ink dried on paper may be sampled by pipetting a small volume of water (e.g., 10uL) onto the dried ink, solubilizing part of the dried ink and retrieving it via the pipette, and then amplifying via PCR. In these examples the resulting liquid is typically diluted substantially, e.g., $>1/1000$, before PCR.

[0231] It is observed via gel electrophoresis of the amplified DNA that the DNA in Ink #4 remains stable after 7 days. Surprisingly, the DNA amplified from the liquid ink of Ink #5 yields a markedly lower DNA concentration compared to Ink #4. In contrast to the liquid ink samples, the DNA in both Ink #4 and Ink #5 used to write on paper on day 0 is observed to be stable at the day 7 timepoint. Notably, the DNA in Ink #4 is further observed to have similar recovery of DNA from both the liquid ink sample and from the sample written on paper. Due to the observed stability, Ink #4 is used for subsequent evaluation.

[0232] The Ink #4 samples are next used in deep sequencing analysis of the NFT-encoding DNA, as summarized in Fig. 42. More specifically, the NFT is encoded into the DNA using 51 nackets, and the heterologous DNA cassette writing method used in the synthesis of the NFT-encoding DNA strands provides a collection of approximately 10^9 unique DNA sequences. PCR analysis of aliquots taken directly from this collection of synthesized DNA sequences yields identification of approximately 10^6 unique DNA sequences (i.e., 1,623,092 unique DNA sequences). This collection of NFT-encoding DNA is incorporated into Ink #4, as above, used in the ink when writing on paper, as above, and subsequently analyzed from the dried ink samples on said paper. Two dried ink samples written on paper are analyzed using PCR and deep sequencing, which are labeled Ink Sample #1 and Ink Sample #2. During analysis, it is observed that Ink Sample #1 has 5,160 unique DNA sequences (1,311 of which are shared with the original DNA sequences identified from the collection previously analyzed) and Ink Sample #2 has 6,218 unique DNA

sequences (2,615 of which are shared with the original DNA sequences identified from the collection previously analyzed). Additionally, Ink Sample #1 and Ink Sample #2 share 442 unique DNA sequences amongst each other. Thus, this shows that the heterologous DNA cassette data writing produces a significant amount of heterogeneity among the DNA sequences, though each DNA strand is ultimately synonymous with all other DNA strands from the same original collection of DNA strands.

[0233] The protocols described above are repeated to further evaluate the stability of the NFT-encoding DNA in ink written on paper over time. More specifically, the DNA in ink written on paper is extracted and analyzed at 2 weeks and 6 weeks post-writing on paper. Notably, the stability at both 2 and 6 weeks are remarkably similar, with no significant difference between time points, as shown in Fig. 43. Additionally, during deep sequencing of the recovered DNA strands, full length nackets of each of the 51 nacket positions were readily identifiable, indicating the absence of any significant breakage in the DNA strands. Moreover, the sequenced nackets yielded consensus sequences for each nacket position, wherein the consensus sequences are useful in the decoding of the DNA sequence back into the original NFT code. By decoding as described herein, e.g., Fig. 34, the original NFT code is reliably recoverable and the object (i.e., ink) is amenable to authentication.

[0234] The protocols described above are further repeated to evaluate the stability of the NFT-encoding DNA in ink written on paper at 8 weeks, as shown in Fig. 44. In this example, 3 replicates (labeled Replicate #1 through Replicate #3) of writing samples are evaluated at 8 weeks post-writing on paper. When comparing nacket analysis, the DNA samples recovered from each of the 3 replicates display remarkable similarity to one another, and are notably similar to the nacket analysis at weeks 2 and 6. After deep sequencing of the 3 replicates of writing samples after 8 weeks, each sample is compared to the other two, with results summarized in Fig. 45. In the first analysis, Replicate #1 is observed to have 8,033 unique nackets, while Replicate #2 is observed to have 9,965 unique nackets. Between Replicate #1 and Replicate #2, 36 nackets are shared. When comparing the number of recovered nackets for each nacket ID, the comparison between Replicate #1 and Replicate #2 yields a linear trend line with $R^2 = 0.954$. In the next analysis, Replicate #2 is observed to have 9,690 unique nackets, while Replicate #3 is observed to have 10,160 unique nackets. Between Replicate #2 and Replicate #3, 311 nackets are shared. When comparing the number of recovered nackets for each nacket ID, the comparison between Replicate #2 and

Replicate #3 yields a linear trend line with $R^2 = 0.969$. In the third analysis, Replicate #1 is observed to have 8,045 unique nackets, while Replicate #3 is observed to have 10,447 unique nackets. Between Replicate #1 and Replicate #3, 24 nackets are shared. When comparing the number of recovered nackets for each nacket ID, the comparison between Replicate #1 and Replicate #3 yields a linear trend line with $R^2 = 0.954$. These data demonstrate, *inter alia*, that between different samples of nacket populations, the majority of synonymous nacket sequences are unique, though a small degree of overlap is possible.

[0235] Following the initial evaluation of DNA stability, heat is used to simulate accelerated aging of DNA sample. In these experiments, a quarter-inch punch of paper with 1 μL of ink is placed into a sealed microcentrifuge tube. The 1 μL of ink is estimated to comprise approximately 4×10^8 molecules of NFT-encoding DNA and 1×10^8 molecules of ddPCR tracer. The sealed microcentrifuge tube containing the ink-marked paper punch is placed in a 75°C oven for various lengths of time before transfer to a 4°C refrigerator for storage before analysis. It is estimated that storing the ink-marked paper at 75°C for 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 days will mimic the room-temperature equivalent of approximately 0, 2.3, 4.6, 6.8, 9.1, 11.4, 13.7, 16.1, 18.3, and 20.5 years, respectively. As a control, an ink-marked paper is stored at -20°C throughout the experiment. After 9 days, wherein each day a sample is moved from the 75°C oven to the 4°C refrigerator, each sample is analyzed using digital PCR. In this case, samples from days 0 and 1 look substantially the same in concentration, while days 2 through 6 each display a steady reduction in DNA concentration after the same number of PCR amplification cycles, and days 7 through 9 display a low concentration of DNA. This likely indicates that the DNA is degrading over time under the accelerated aging conditions at 75°C , though the extent of degradation is unclear. Next, the aged samples are amplified via PCR at varying cycle numbers to yield sufficient material for sequencing. In this case, the ddPCR tracer added to the NFT-encoding DNA in the ink marking the paper punch is used to amplify a 700 bp length of DNA. While quantifying the amplified DNA, it is observed that approximately 6.5% of the DNA is recovered in the day 0 sample. Next, approximately 3% of the DNA is recovered in the day 1 (approx. 2.3 year equivalence) sample, approximately 1% of the DNA is recovered in the day 2 (approx. 4.6 year equivalence) sample, and progressively less DNA is recovered in each subsequently aged sample. The results are displays in Fig. 46, wherein a logarithmic decline in DNA recovery is observed.

[0236] Continuing the PCR analysis of the DNA after accelerated aging, amplicons on each end of the 700 bp length of DNA targeted by the ddPCR tracer allow for analysis of double-stranded DNA breakage in the aged samples. Surprisingly, the DNA stays largely resistant to breakage throughout the evaluated time points, with less than 10% breakage observed for days 0, 1, and 2 (approx. 0, 2.3, and 4.6 year equivalence), while days 3, 4, and 5 (approx. 6.8, 9.1, and 11.4 year equivalence) display 10-25% breakage. However, days 6 through 9 display more notable DNA breakage, between 40-65% breakage. These results are summarized in Fig. 47.

[0237] Lastly, by directly comparing the sequenced DNA samples after undergoing accelerated aging, it is observed that the error rate of the DNA only slightly increases over time, while the sequence efficiency (i.e., proportion of DNA that are “correct” reads or consensus sequences) decreases over time. These results are summarized in Fig. 48. This is emphasized by the sequence length distribution of Fig. 49, wherein the sequence length shifts over time from a single prominent length of DNA to a series of shorter DNA strands. Thus, these results indicate that the DNA does sustain damage over time, but the error rate in the DNA sequence remains relatively stable and the DNA is still capable of decoding and recovery of consensus sequences, even after an equivalence of 20 years accelerated aging.

EXAMPLE 2: ENCAPSULATION AND EXTRACTION OF DNA FROM SILICA BEADS

[0238] It is known that DNA can be encapsulated in nanometer silica beads, which can be fused into various materials that are used to print or cast objects in any shape and subsequently recovered. See, e.g., Koch J, et al., “A DNA-of-things storage architecture to create materials with embedded memory.” Nat. Biotechnol. (2020)38(1):39-43; e.g., U.S. Patent No. 9,850,531, “Molecular code systems”; e.g., Bossert, et al., “A hydrofluoric acid-free method to dissolve and quantify silica nanoparticles in aqueous and solid matrices” Sci. Rep. (2019)9:7938, the contents of each of which are incorporated herein by reference.

[0239] For example, a machine-readable code is converted into a collection of DNA strands using heterologous DNA cassette data writing, as described in Example 1. After synthesis, but before incorporating the DNA into a material or object, the DNA is encapsulated into silica beads, e.g., silica microspheres.

[0240] Silica seed particles are mixed with a solution of the free DNA encoding the NFT, which coats the seed particles with DNA strands. Optionally, the silica seed particles may be modified

with amine-bearing functional groups to allow for enhanced interaction with DNA polymers. The DNA-coated seed particles are subsequently mixed with a solution of tetra ethoxy silane (TEOS) and base in ethanol to grow a SiO₂ layer around the DNA, yielding the silica beads with DNA encapsulated therein. More specifically, 5 µL of free DNA (at 28 ng/µL) is mixed with 10 µL of silica seed particles (at 60 mg/mL) in 500 µL TE buffer. The resulting mixture is centrifuged (at 21,500 g) for 1 minute, the supernatant is removed, and the pellet is dispersed in 1 mL ethanol. To this suspension, 2 µL APTES is added with 20 µL TEOS and 20 µL TE buffer. The solution is allowed to react overnight at room temperature while shaking, after which the solution is again centrifuged and the precipitate is washed with ethanol and TE buffer before re-suspension.

[0241] Following encapsulation, a first extraction protocol is used. In this first extraction protocol, the DNA-encapsulating silica beads are dissolved in buffered oxide etch solution, wherein the oxide etch solution comprises an aqueous mixture of ammonium fluoride and hydrofluoric acid, which may be done in 0-50°C, though readily proceeds at room temperature. The beads readily dissolve within several seconds in the oxide etch solution, yielding the original free DNA within a high-salt solution (e.g., F⁻, NH₄⁺, and SiF₆²⁻), though it is thought that the relatively high pKa of hydrofluoric acid prevents damage to the DNA. More specifically, 5 µL of silica beads encapsulating DNA is added to 10 µL of a buffer oxide etch solution (0.34g NH₄F and 10g HF (at 1%) in TE buffer), and shaken for 1 minute. The mixture transitions from a turbid to clear solution, and the resulting solution is dialyzed against 10 mL of water for 30 minutes. Following dialysis, the free DNA is analyzed via PCR, as described in Example 1.

[0242] A second extraction protocol is also useful as an alternative, particularly since the use of hydrofluoric acid is often undesirable. In this alternative extraction protocol, the etch solution used for dissolving the silica beads is composed of aqueous potassium hydroxide. In this case, 10 µg/mL of silica beads is mixed with 1M KOH in an aqueous solution with a pH of 12, wherein the silica beads dissolve overnight at room temperature. Alternatively, 10 µg/mL of silica beads is mixed with 0.1M KOH in an aqueous solution with a pH of 12, wherein the silica beads dissolve within 15 minutes under 1500 W of microwave radiation. Following silica bead dissolution and extraction of the encapsulated DNA, the free DNA is dialyzed and analyzed as described above.

[0243] Although the above example(s) have been described using exemplary procedures, materials, objects, concentrations, or processes for implementing the present disclosure, it should

be understood by those skilled in the art that alternative procedures, materials, objects, concentrations, or processes or other combinations and sequences of the procedures, materials, objects, concentrations, and processes described herein may be used or performed that achieve the same function(s) and result(s) described herein and which are included within the scope of the present disclosure. For example, beyond the examples and embodiments described above, additional exemplary embodiments have been developed with success, including applications of the present invention in latex paint (both free and encapsulated DNA), acrylic paint (both free and encapsulated DNA), industrial inkjet printer ink (free DNA), perfume (free DNA), oil paint (encapsulated DNA), permanent marker ink (free and encapsulated DNA), stamp-pad ink (free DNA), watercolor paint (free DNA), and 3D printing plastic (encapsulated DNA).

CLAIMS

What is claimed is:

1. A population of deoxyribonucleic acid (DNA) sequences encoding data useful in the authentication of objects and for protection against counterfeiting (e.g., selected from DNA 1, et seq. and/or DNA 2, et seq.), comprising nucleic acid data packets ("nackets"), wherein each nacket is encoded by a plurality of DNA molecules encoding the same data, wherein the sequences of the DNA molecules are heterogeneous.
2. The population of DNA sequences of claim 1, wherein the DNA sequences are prepared using heterologous cassette data writing, wherein two or more cassette sequences are provided for a single bit or combination of bits in a machine-readable code, such that all or nearly all of the DNA molecules in the nacket encode the same data, but the sequences of the individual molecules exhibit extremely high variation, wherein the nackets comprise a plurality of heterologous cassettes.
3. The population of DNA sequences of claim 1 or 2, wherein the data is in n -bit code wherein n is greater than 1, e.g., binary or ternary code.
4. The population of DNA sequences of any foregoing claim, wherein the DNA sequences are prepared from heterologous cassettes encoding the same bit or bits of data, wherein the percent abundance of the different cassette variants used in writing the DNA provides a unique and distinguishable feature of the DNA.
5. The population of DNA sequences of any foregoing claim, wherein the data encoded in the DNA is a nonfungible token (NFT).
6. The population of DNA sequences of any foregoing claim, wherein the one or more DNA sequences and/or cassettes contain one or more topoisomerase recognition sequences,

e.g., wherein the topoisomerase recognition sequence is 5'-CCCTT-3', 5'-TCCTT-3', 5'-CCCTG-3', or 5'-TGACT-3'.

7. The population of DNA sequences of any foregoing claim, wherein the DNA comprises cassettes, wherein each cassette comprises (i) an information domain having sequence which corresponds to one or more bits in a machine-readable code, and (ii) a topoisomerase recognition sequence, wherein the cassette is 18-25 nucleotides in length.
8. The population of DNA sequences of any foregoing claim, wherein the DNA is incorporated into or associated with goods for purposes of identifying and authenticating the goods.
9. The population of DNA sequences of any foregoing claim, wherein the DNA is adsorbed onto, incorporated into, or encapsulated by silica beads or particles.
10. A method of object authentication (e.g., according to any of Method 1, et seq., *supra*), comprising:
 - i. synthesizing a population of DNA sequences, e.g., according to claim 1, comprising nucleic acid data packets ("nackets"), wherein each nacket contains a plurality of DNA molecules encoding the same data, wherein the sequences of the DNA molecules are heterogeneous;
 - ii. incorporating said DNA sequences into or onto an object;
 - iii. extracting said DNA sequences from the object; and
 - iv. analyzing the extracted DNA sequences;
 - v. optionally, comparing the analyzed DNA sequences to a database of DNA sequences;
 - vi. optionally, confirming object authenticity.
11. The method of claim 10 wherein the DNA sequences are synthesized by sequential addition of DNA cassettes to DNA receptor strands, wherein in each sequential addition step the cassettes comprise a heterologous population of synonymous cassettes, such that the

cassettes have at least two different sequences encoding the same data in a machine-readable code (e.g., binary or ternary code).

12. The method of claim 10 or 11, wherein the cassettes are conjugated together using a ligase enzyme.
13. The method of claim 10 or 11, wherein the cassettes are conjugated together using a topoisomerase enzyme.
14. The method of any of claims 10 to 13, wherein the DNA sequences comprise DNA sequences synthesized using a transferase-based synthesis and data encoding.
15. The method of any previous claim wherein the cassettes are synthesized by sequential addition of cassettes to DNA receptor strands using an inkjet printing head (e.g., a piezoelectric print head), wherein each cassette comprises multiple nucleotides, wherein in each sequential addition step the cassettes comprise a heterogeneous population of cassettes of at least two different sequences encoding the same data in a machine-readable code (e.g., binary or ternary code), and wherein the cassettes are dispensed by an inkjet writing print head on at least one writing spot on a wafer array, the head or nozzle writing the same code to a plurality of polymer memory strands dispensed on the at least one spot, e.g., comprising the following steps:
 - a) loading the desired spot to be written with a starter polymer or DNA attached at one end to the desired spot;
 - b) washing the surface of the spot;
 - c) positioning an inkjet nozzle having a heterogeneous population of cassettes wherein the population comprises cassettes having at least two different sequences, but all encoding the same information in one or more bits (e.g., 1 or 0, or 00, 01, 10, 11, etc. in binary code) over the desired spot to be written corresponding to the unique code;
 - d) causing the inkjet nozzle to release a droplet comprising the heterogeneous population of cassettes onto the spot, thereby writing a bit or portion of the unique code to the DNA or polymer memory strings (or strands) associated with the spot; and

e) washing the surface of the spot;

optionally further comprising steps f) – i):

f) causing the inkjet nozzle to release a droplet of deblock/adaptor reagent onto the spot;

g) washing the surface of the spot;

h) repeating steps (c) through (g) until the unique code has been written in the memory string at the spot; and

i) removing the memory strings from the spot and flowing the memory strings from the spot into a collection or storage container for later incorporation into or onto an object.

16. The method of claim 15, wherein the cassettes are added by topoisomerase mediated ligation; for example, by:

(i) reacting double-stranded acceptor DNA strands with topoisomerases charged with double-stranded DNA cassettes from the heterologous population of cassettes covalently bound to the topoisomerases,

wherein a strand of the acceptor DNA has a 5' overhang,

wherein each cassette comprises an informational sequence, a topoisomerase recognition sequence, and 5' overhangs on both strands,

wherein the 5' overhang of the strand of the oligomer that does not bear the topoisomerase ("bottom strand") is complementary to the 5' overhang of the acceptor DNA but is not complementary to the 5' overhang of the strand bearing the topoisomerase ("top strand") of the cassette,

wherein the 5' end of the strand bearing the topoisomerase ("top strand") of the cassette and 5' end of the acceptor DNA are not protected, e.g., not phosphorylated (i.e., 5'-OH), and

wherein the topoisomerase charged with a double-stranded DNA cassette is delivered to the location of the acceptor strand by a piczo-electric inkjet nozzle;

(ii) reacting the acceptor DNA thus extended in step (i) with a topoisomerase charged with a further double-stranded DNA cassette,

wherein the further cassette comprises an informational sequence that is the same as or is different from any informational sequence in the cassette of step (i), a topoisomerase recognition sequence, and 5' overhangs on both strands,

wherein the 5' overhang of the strand of the further cassette not bearing the topoisomerase ("bottom strand") is complementary to the 5' overhang of the extended acceptor DNA but is not complementary to the 5' overhang of the strand of the further cassette bearing the topoisomerase ("top strand"), and

wherein the 5' end of the strand bearing the topoisomerase ("top strand") of the further cassette is not protected, e.g., not phosphorylated (i.e., 5'-OH); and

(iii) repeating steps (i) and (ii) until the desired nucleotide sequence is obtained; wherein there is optionally a washing step after step (i) and/or after step (ii); and

optionally, wherein the desired nucleotide sequence thus obtained is further reacted with a terminal sequence comprising one or more replication primers, such as one or more PCR primer sequences.

17. A method (e.g., according to any of Method 2, et seq., *supra.*) of any preceding claim wherein the nackets or cassettes used to make the nackets comprise a desired code, e.g., a ternary code, using a DNA or polymer strand or memory string, wherein the data is encoded in a series of transitions between non-identical nucleotides, with one bit for each such transition, comprising:

i. providing a reaction mixture comprising one or more transferase enzyme, e.g., terminal deoxynucleotidyl transferase (TdT) and one or more dNTP degrading enzyme, e.g., apyrase;

ii adding to the reaction mixture deoxyribonucleotide triphosphates (dNTPs), e.g., selected from dATP, dCTP, dGTP, and dTTP;

iii. waiting until the dNTPs of step (ii) are added or degraded;

iv. repeating steps (ii) and (iii) until the desired bit sequence is reached, wherein non-identical dNTP species are used in any two consecutive additions

thereby providing a population of DNA molecules encoding the desired data string.

18. A method of object authentication (e.g., according to any of Method 3, et seq., *supra*), comprising:
- i. synthesizing one or more DNA sequences, e.g., according to claim 1, comprising nucleic acid data packets (“nackets”), wherein each nacket contains a plurality of DNA molecules encoding the same data, wherein the sequences of the DNA molecules are synthesized using one or more transferase enzymes, e.g., according to any of DNA 2, et seq., *supra*;
 - ii. incorporating said one or more DNA sequences into or onto an object;
 - iii. extracting said one or more DNA sequences from the object; and
 - iv. analyzing the extracted one or more DNA sequences;
 - v. optionally, comparing the analyzed one or more DNA sequences to a database of DNA sequences;
 - vi. optionally, confirming object authenticity.
19. A method for writing an attack resistant digital code using DNA, comprising:
- i. receiving a desired digital code to be written, the desired code being grouped into four two-bit binary codes to be written (e.g., 00, 01, 10, 11);
 - ii. providing four predetermined mixtures of a predetermined number of unique DNA cassette strings, each mixture corresponding to a different predetermined two-bit binary code value, each mixture having a predetermined proportion of the unique DNA cassettes within the mixture, and the unique DNA cassette strings of each mixture being different from the DNA cassette strings in the other mixtures;
 - iii. depositing a droplet of the mixture associated with a given two-bit binary code to be written onto a substrate to add a DNA cassette string to an encoded DNA string being written, the droplet comprising the predetermined mixture of the unique cassettes associated with the given two-bit binary code; and
 - iv. repeating the depositing until the desired code is written onto the encoded DNA string.
20. The method of claim 19, further comprising, after the desired code is written, adding an end cap to the encoded DNA string.

21. The method of claim 20, wherein the end cap contains information about the desired digital code or how to read the code.
22. The method of any of claims 19 to 21, wherein the substrate has an acceptor DNA strand having one end attached to the substrate and an opposite end being available to attach to one of the unique DNA cassettes to be added.
23. The method of any of claims 19 to 22, wherein the predetermined number of unique DNA cassettes for one of the mixtures is different from at least one other of the mixtures.
24. The method of any of claims 19 to 23, wherein the desired digital code is encoded in an NFT with authentication data and stored on a blockchain.
25. The method of any of claims 19 to 24, wherein the encoded DNA string is embedded in a physical object to be authenticated.
26. A method for writing an attack resistant digital code using DNA, comprising:
- i. receiving a desired digital code to be written, the desired code being grouped into a plurality of n -bit binary codes to be written, where n is greater than 1;
 - ii. providing at least two predetermined mixtures of a predetermined number of unique DNA cassette strings, each mixture corresponding to a different predetermined n -bit binary code value, each mixture having a predetermined proportion of the unique DNA cassettes within the mixture, and the unique DNA cassette strings of each mixture being different from the DNA cassette strings in the other mixtures;
 - iii. depositing a droplet of the mixture associated with a given n -bit binary code to be written onto a substrate to add a DNA cassette string to an encoded DNA string being written, the droplet comprising the predetermined mixture of the unique cassettes associated with the given n -bit binary code; and
 - iv. repeating the depositing until the desired code is written onto the encoded DNA string.

27. The method of claim 26, further comprising, after the desired code is written, adding an end cap to the encoded DNA string.
28. The method of claim 27, wherein the end cap contains information about the desired digital code or how to read the code.
29. The method of any of claims 26 to 28, wherein the substrate has an acceptor DNA strand having one end attached to the substrate and an opposite end being available to attach to one of the unique DNA cassettes to be added.
30. The method of any of claims 26 to 29, wherein the predetermined number of unique DNA cassettes for one of the mixtures is different from at least one other of the mixtures.
31. The method of any of claims 26 to 30, wherein the desired digital code is encoded in an NFT with authentication data and stored on a blockchain.
32. The method of any of claims 26 to 31, wherein the encoded DNA string is embedded in a physical object to be authenticated.
33. A method for writing an attack resistant digital code using DNA, comprising:
- i. receiving a desired digital code to be written, the desired digital code being grouped into four two-bit binary codes to be written (e.g., 00, 01, 10, 11);
 - ii. providing four sets of unique DNA cassette strings, each set comprising a predetermined number of unique DNA cassettes and each set corresponding to a different predetermined two-bit binary code value, such that each set of unique cassettes corresponding to different two-bit binary code and each set of unique cassette strings being different from the other DNA cassette strings;
 - iii. randomly selecting one of the unique cassettes corresponding to a given two-bit binary code to be written, as a selected unique cassette;

- iv. depositing a droplet of the selected unique cassette associated with the given two-bit binary code to be written onto a substrate to add the selected unique cassette to an encoded DNA string being written;
- v. repeating the selecting and depositing until the desired code is written onto the encoded DNA string on a given writing spot on the substrate; and
- vi. counting the number of times each unique cassette is used for each two-bit binary code written.

34. A method for writing an attack resistant digital code using DNA, comprising:

- i. receiving a desired digital code to be written, the desired digital code being grouped into a plurality of n -bit binary codes to be written, where n is greater than 1;
- ii. providing at least two sets of unique DNA cassette strings, each set comprising a predetermined number of unique DNA cassettes and each set corresponding to a different predetermined n -bit binary code value, such that each set of unique cassettes corresponding to different n -bit binary code and each set of unique cassette strings being different from the other DNA cassette strings;
- iii. randomly selecting one of the unique cassettes corresponding to a given n -bit binary code to be written, as a selected unique cassette;
- iv. depositing a droplet of the selected unique cassette associated with the given n -bit binary code to be written onto a substrate to add the selected unique cassette to an encoded DNA string being written;
- v. repeating the selecting and depositing until the desired code is written onto the encoded DNA string on a given writing spot on the substrate; and
- vi. counting the number of times each unique cassette is used for each n -bit binary code written.