

АНАЛІЗ МОЖЛИВОСТЕЙ ДОВГОСТРОКОВОГО ЗБЕРІГАННЯ ДАНИХ НА ДНК

А. А. Крючин, Є. В. Беляк, Є. А. Крючина¹, О. В. Шиховець

Інститут проблем реєстрації інформації НАН України

¹Київська міська клінічна лікарня № 10

Наведено результати критичного аналізу перспектив застосування ДНК пам'яті для довгострокового зберігання інформації. Показано, що ДНК пам'ять забезпечує як запис інформації з високою щільністю запису, так і її довгострокове зберігання. Наведені дані свідчать, що технології ДНК пам'яті зможуть сприяти суттєвим змінам в архівному зберіганні даних. Проведено аналіз галузей застосування ДНК пам'яті. Виконано дослідження кодів, що використовуються для представлення даних у ДНК пам'яті. Детально проаналізовано декілька підходів щодо проектування кодонів ДНК і різноманітних підходів до зберігання даних, визначено плюси та мінуси кожного з підходів. Обговорюються методи стеганографії з використанням молекул ДНК для безпечного зберігання даних.

Ключові слова: ДНК пам'ять, секвенування, полімеразна ланцюгова реакція, концепція «ДНК речей».

ANALYSIS OF THE POSSIBILITIES OF LONG-TERM STORAGE OF DATA ON DNA

A. A. Kryuchyn, Ie. V. Belyak, Ye. A. Kryuchyna¹, A. V. Shikhovets

Institute for Information Recording of the National Academy of Sciences of Ukraine

¹Kyiv City Clinical Hospital № 10

Background. For several decades, progress in storage technologies has been measured primarily in terms of storage capacity and data read / write speed, with the amount of information generated increasing every day. The purpose of the study was to analyze the various technologies used to store digital data on DNA, the pros and cons of each of the methods used, as well as the advantages and limitations of using DNA as an information carrier.

Materials and methods. Results. The reliance on DNA as a dense storage medium with high storage capacity and its ability to withstand extreme environmental conditions has increased over the past few years. The presented work presents the results of a critical analysis of the prospects for using DNA memory for long-term storage of information. It is shown that DNA memory provides both information recording with a high recording density and its long-term storage. These data indicate that DNA memory technologies can contribute to significant changes in the archival storage of data. The analysis of the areas of application of DNA memory, codes used to represent data in DNA memory. A number of approaches to the design of DNA codons and data storage are analyzed in detail, certain pros and cons of each of them.

Conclusions. Methods of steganography using DNA molecules for the safe storage of information are discussed. At the same time, the high cost of DNA, the impossibility of quick access to information, the need for expensive equipment for reading and writing makes this type of storage device still suitable for limited use.

Key words: DNA memory, sequencing, polymerase chain reaction, concept of «DNA of things».

АНАЛИЗ ВОЗМОЖНОСТЕЙ ДОЛГОСРОЧНОГО ХРАНЕНИЯ ДАННЫХ НА ДНК

А. А. Крючин, Е. В. Беляк, Е. А. Крючина¹, А. В. Шиховець

Інститут проблем реєстрації інформації НАН України

¹Киевская городская клиническая больница № 10

В течение последних нескольких десятилетий прогресс в технологиях хранения информации измерялся, прежде всего, с точки зрения емкости накопителей и скорости чтения/записи данных, при этом объем генерируемой информации увеличивается с каждым днем. Надежды на ДНК как на плотный носитель информации с большой емкостью и ее способность выдерживать экстремальные условия окружающей среды существенно возросли за последние несколько лет. В представленной работе приведены результаты критического анализа перспектив применения ДНК памяти для долгосрочного хранения информации. Показано, что ДНК память обеспечивает как запись информации с высокой плотностью записи, так и ее долгосрочное хранение. Приведенные данные свидетельствуют, что технологии ДНК памяти смогут способствовать существенным изменениям в архивном хранении данных. Проведен анализ областей применения ДНК памяти, кодов, используемых для представления данных в ДНК памяти. Детально проанализированы ряд подходов к проектированию кодонів ДНК и хранению данных, определенные плюсы и минусы каждого из них. Обсуждаются методы стеганографии с использованием молекул ДНК для безопасного хранения информации. В тоже время показано, что дороговизна ДНК, невозможность быстрого доступа к информации, потребность в дорогостоящем оборудовании для считывания и записи делает этот тип накопителя пока пригодным для ограниченного использования.

Ключевые слова: ДНК память, секвенирование, полимеразная цепная реакция, концепция «ДНК вещей».

Вступ. РОбсяги цифрової інформації, що підлягають зберіганню, швидко зростають і випереджають можливості традиційних технологій зберігання даних [1-6]. При організації зберігання даних необхідно забезпечити не тільки запис великих обсягів інформації, але й їх довгострокове зберігання. Існуючі носії інформації не можуть повною мірою забезпечити вирішення цього завдання [1, 6]. Так, жорсткі та оптичні диски, flash-пам'ять тощо можуть забезпечити збереження інформації лише протягом 50-100 років. Одним із альтернативних методів довгострокового зберігання інформації є використання молекул ДНК (дезоксірибонуклеїнової кислоти) в якості інформаційних носіїв. Завдяки своїй довговічності та величезній інформаційній щільності ДНК, молекула, що кодує біологічну інформацію, стала перспективним архівним носієм інформації. В останні роки відбулися розробки щодо читання та написання різних форм даних про ДНК, кодів для шифрування даних і використання ДНК як способу «таємного письма» — стеганографії та криптографії. Однак, суттєвим недоліком цього методу є досить високий рівень помилок, пов'язаних із недосконалістю використовуваних методів і часовою хімічною деградацією молекул, що виникають при визначенні послідовності (секвенуванні) генетичного коду, що міститься в ДНК.

Мета роботи: проаналізувати різні технології, використані для зберігання цифрових даних на ДНК, плюси та мінуси кожного із застосованих методів, а також переваги та обмеження використання ДНК як носія інформації.

Матеріали та методи дослідження. Проведено теоретичний аналіз і узагальнення, систематизація результатів досліджень відповідно до рекомендацій PRISMA Group із використанням таких провідних наукометричних баз даних як ScienceDirect, PubMed, Emerald, IEEE Xplore, Taylor & Francis за період з 2003 по 2020 рр. за ключовими словами DNA data storage, DNA steganography, Sequencing, у тому числі в українському та російському перекладах.

Результати та їх обговорення. Перспективним напрямом довгострокового зберігання великих обсягів інформації є використання ДНК пам'яті [3, 5-10]. ДНК є середовищем для запису даних із високими щільністю запису та часовою стабільністю. Розвиток нових технологій як у синтезі ДНК, так і в секвенуванні, робить ДНК все більш доступним середовищем для зберігання даних [5, 6]. Системи зберігання, побудовані на основі синтетичних ДНК, зможуть обійти два ключових

обмеження нинішніх систем — обмежений строк життя носіїв і низьку щільність зберігання інформації. Носії на основі ДНК зможуть зберігати інформацію тисячі років, а в пристрої вагою один грам може поміститися 1021 Байт цифрових даних. Верхня межа щільності зберігання інформації на ДНК була визначена приблизно як $4,606 \times 10^{20}$ Байт/г [5], але більш практичним показником є об'ємна щільність. У дослідженні Church G. et al. [11] об'ємна щільність молекул ДНК була апроксимована щільністю чистої води, що дало щільність інформації $4,606 \times 10^{17}$ Байт/мм³. Сьогодні більшість даних зберігаються в пристроях, що мають обмежений термін придатності, тому періодично їх доводиться переписувати на нові носії. ДНК — це перспективний засіб зберігання, тому що строк його експлуатації становить тисячі років [5, 6]. Області застосування технології зберігання даних на ДНК постійно розширюються і цьому сприяють чисельні розробки з удосконалення технології зберігання даних на ДНК. У напрямку створення ДНК пам'яті досягнуто певних успіхів. Так, фахівцями Microsoft і Вашингтонського університету розроблено автоматизовану систему ДНК зберігання даних. Спеціальне програмне забезпечення перетворює одиниці та нулі цифрових даних у послідовності нуклеотидів (аденін (А), тимін (Т), гуанін (G) і цитозин (С)), що представляють собою «будівельні блоки» ДНК. Потім система використовує стандартне лабораторне обладнання для подачі необхідних рідин і реагентів у синтезатор, що збирає виготовлені фрагменти ДНК і поміщає їх в ємність для зберігання [15]. В лютому 2018 р. була створена «бібліотека праймерів» для організації довільного доступу до ДНК. Кожен із праймерів «прив'язаний» до конкретного ланцюжку, тому за допомогою полімеразної ланцюгової реакції (ПЛР) можна обрати будь-який із них і отримати доступ до записаних даних [7]. У червні 2019 р. усі 16×10^9 Байт тексту з англійської версії Вікіпедії були закодовані в синтетичну ДНК [13]. Розроблено оригінальний метод зчитування у ДНК пам'яті з меншою кількістю операцій секвенування, необхідних для отримання конкретних даних у бібліотеці ДНК пам'яті, з використанням методу вдалося відтворити 35 файлів (із загальним розміром $> 200 \times 10^6$ Байт) без помилок. Метод може бути поширений на бібліотеку ДНК пам'яті з кількома терабайтами ємності. Розроблений метод використано у спробі зберегти найбільший на сьогодні обсяг даних у ДНК [14].

Зберігання даних на ДНК дозволило створити оригінальні технології зберігання даних, одна з них - це концепція «ДНК речей» (DoT) [15]. Технологія DoT передбачає кодування цифрових даних у молекули ДНК, які потім вбудовуються в об'єкти. На відміну від Інтернету речей (IoT), який являє собою глобальну мережа підключених до Інтернету фізичних пристроїв - «речей» оснащених сенсорами, датчиками і пристроями передачі інформації, DoT створює об'єкти, які є незалежними. DoT представляє засіб зберігання великих обсягів інформації практично про будь-які об'єкти. За допомогою цього методу можна інтегрувати інформацію для 3D-друку в об'єкт, так що

через десятиліття або навіть століття можна буде отримати цю інформацію безпосередньо від самого об'єкту. Технології DoT в молекули ДНК записують дані, а потім ці молекули розміщують у нанорозмірних гранулах з кварцового скла, які вбудовують в різні матеріали, що використовуються для друку або відливання об'єктів довільної форми. Для підтвердження концепції DoT був надрукований на 3D-принтері Стенфордський кролик, який містив для його синтезу дані на ДНК об'ємом 45 Кбайт. Було синтезовано п'ять поколінь кроликів, кожне з використанням даних попереднього покоління, без додаткового синтезу ДНК. Схема процесу DoT і пробного 3D-друку наведені на рис.1 [15].

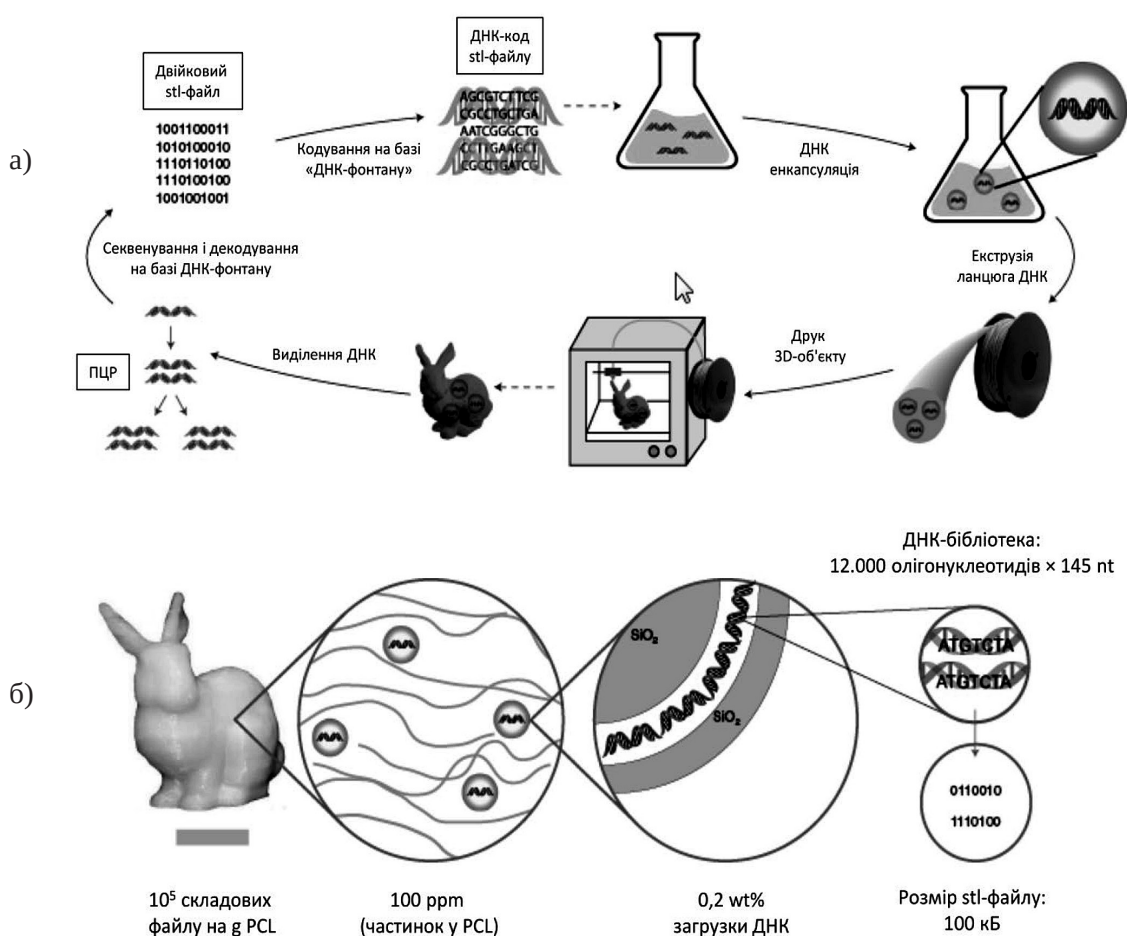


Рис. 1 Схема процесу DoT і пробного 3D-друку Стенфордського кролика: а) етапи процесу DoT - створення цифрового файлу, кодування в бібліотеку олігонуклеотидів ДНК з використанням фонтанного кодування (Fountain coding), інкапсуляція синтезованої бібліотеки ДНК методом золь-гель-синтезу в маленькі скляні частки і змішування з полікапролактоном (PCL), який екструдується в стандартну нитку для 3D-друку, 3D друк за допомогою нитки PCL, що містить ДНК, об'єкту, який визначено у цифровому файлі. Бібліотеку ДНК можна витягти з будь-якої частини надрукованого об'єкту і утворити додаткові копії за допомогою ПЛР. За допомогою секвенування ДНК і розшифровки закодованої інформації можна отримати вихідний файл у форматі stl для 3D-друку нових об'єктів; б) схема процесу 3D друку Стенфордського кролика, який містить інформацію, необхідну для повторного друку об'єкта [15]

DoT може застосовуватися для зберігання електронних медичних карт у медичних імплантах, для стеганографічних цілей і створення об'єктів, що містять дані про об'єкт інтегровані в матеріал самого об'єкту [17]. Можливість DoT використовуватися для стеганографічних цілей була продемонстрована шляхом створення прозорих лінз, які містять відео, інтегроване в матеріал лінз [15,16].

Базові технології створення ДНК пам'яті. ДНК пам'ять може забезпечити технологічний стрибок в області зберігання даних завдяки високій щільності зберігання, довговічності та енергоефективності. Загальна система зберігання інформації на основі ДНК показана на рис. 2. [2].

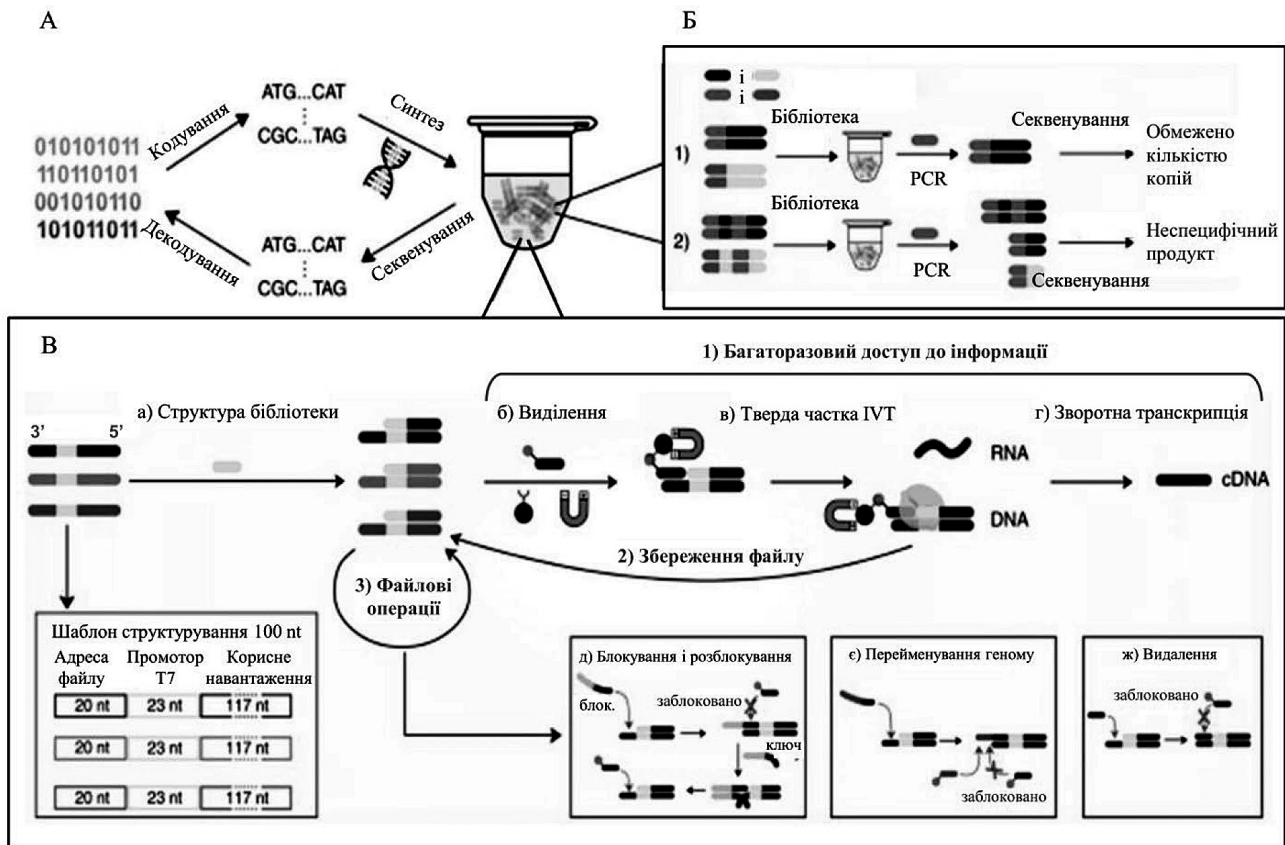


Рис. 2. Загальна структура системи зберігання даних на основі ДНК і організація доступу в ДНК пам'ять

На рис. 2 бачимо такі складові загальної структури системи зберігання даних на основі ДНК і організації доступу в ДНК пам'ять: А) загальна структура системи зберігання даних на основі ДНК, що включає кодування цифрової інформації в нуклеотидні послідовності, синтез, зберігання, секвенування ДНК і декодування інформації; Б) організація доступу до файлів на основі ПЛР; В) схема DORIS (динамічні операції і багаторазове сховище інформації). Зв'язки ss-ds ДНК забезпечують доступ до інформації, що повторюється, транскрипції *in vitro*, зворотної транскрипції і повернення файлів в базу даних. Зв'язки ss-ds ДНК дозволяють виконувати операції з файлами в сховище, включаючи блокування, розблокування, перейменування та видалення [2].

Для підвищення ефективності роботи ДНК пам'яті тривають дослідження щодо вирішення проблем, пов'язаних з кодуванням цифрової інформації в нуклеотидні послідовності, синтезом, зберіганням та секвенуванням ДНК, декодуванням інформації [2,5,6]. На сьогодні виконуються роботи щодо технологій доступу та управління підмножинами даних в базах даних, які зберігаються в ДНК, особливо динамічних. Синтез та секвенування ДНК є значно повільнішими процесами, ніж електронний запис та читання даних, тому ДНК пам'ять буде використовуватися головним чином для архівування або "холодного" зберігання даних. Архівна база даних на ДНК, не обов'язково повинна бути повністю статичною і її динамічні властивості будуть поступово покращуватися.

Реалізація динамічних властивостей наблизить системи зберігання на основі ДНК до практичної реалізації [2]. У порівнянні з іншими формами зберігання даних на молекулярному або атомарному рівнях, пам'ять на ДНК унікальна через простоту копіювання (наприклад, за допомогою ПЛР). Технології високопродуктивного секвенування та синтезу ДНК еволюціонували та зробили зберігання інформації у синтетичній ДНК все більш реалістичною альтернативою традиційним методам тривалого зберігання даних. Однак, було виявлено, що у процесі секвенування можливо виникнення помилок розпізнавання олигонуклеотидів, що вимагає застосування сучасних кодів корекції помилок, здатних виправляти випадання певних послідовностей [18].

Щоб зберігання даних на синтетичних ДНК стало альтернативою архівуванню на традиційних носіях, необхідно фізично зберігати в одному місті багато унікальних послідовностей ДНК, а потім організувати довільний та надійний доступ до даних. Випадковий доступ вимагає набагато менш ресурсів для відновлення даних, оскільки лише відповідні файли послідовно відтворюються та аналізуються. Теоретично для досягнення максимальної щільності запису при виконання довільного доступу ПЛР потрібна лише одна копія кожної послідовності. Однак, це не так з двох причин: стохастичні варіації числа копій, які виникають внаслідок вибіркової сукупності пулу під час довільного доступу, та варіації числа копій, які виникають внаслідок синтезу. Визначення мінімального числа копій для кожної реакції ПЛР має вирішальне значення для зберігання

даних ДНК, без нього можна зберігати занадто мало копій для доступу до даних або занадто багато, марно витрачаючи ємність носія [14]. Для організації випадкового доступу запропоновано розміщувати послідовності файлів ДНК у герметичні капсули з оксиду кремнію, які поверхнево марковані одноланцюговими штрих-кодами ДНК. Штрих-коди обрані для представлення метаданих файлів, що забезпечує ефективний та прямий вибір наборів файлів [19].

Щоб конкурувати з сучасними технологіями запису даних за швидкістю відтворення та запису інформації необхідно зменшити вартість технології синтезу на 7–8 порядків та 6 порядків вартість технології секвенування ДНК [20]. Вартість зберігання даних на ДНК поки що значно перевищує вартість зберігання на магнітних носіях. Результати проведених досліджень довели, що основним фактором, який визначає витрати на зберігання даних на ДНК, є синтез, і якщо вартість синтезу зменшиться на 2 порядки (порівняно з цінами 2013 р.), вартість зберігання даних на ДНК протягом десяти років буде менше вартості зберігання даних на магнітних носіях [21]. У дослідженні Erlich Y. et al. [22] наведені дані про меншу вартість зберігання даних на ДНК, яка становила \$ 3500 за МБайт, але вона була розрахована при використанні більш економічного методу для синтезу ДНК, який передбачав використання більш ефективного алгоритму виправлення помилок. Програмою молекулярного зберігання інформації (прийнятою у США) прогнозується, що до 2023 р. вартість синтезу ДНК зменшиться до 10-10 дол. /bp (п.о.) [6]. Оцінки вартості синтезу ДНК для зберігання даних в динаміці наведено на рис.3. [6]

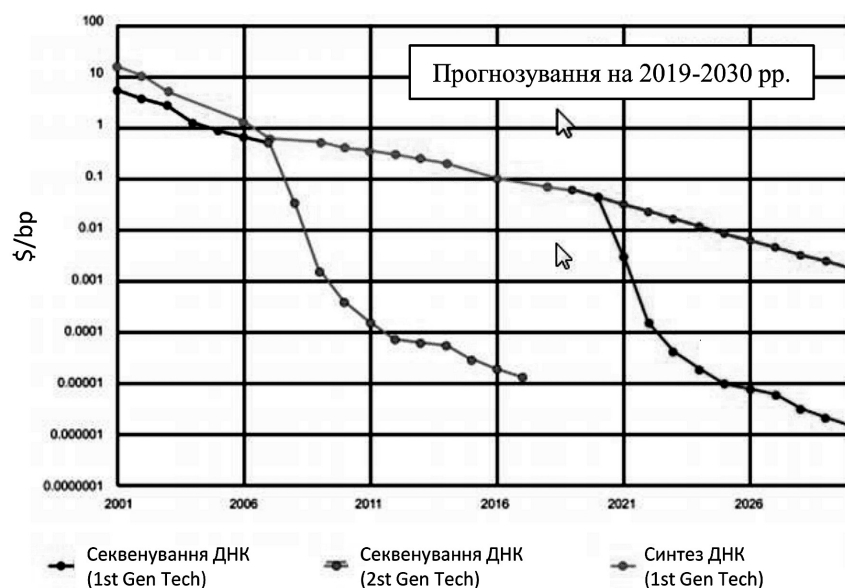


Рис. 3. Оцінки вартості синтезу ДНК для зберігання даних [6]

На сьогодні більшість спроб створення систем зберігання даних на ДНК було зроблено *in vitro*. Однак ДНК живих клітин можуть стати ідеальним середовищем для зберігання інформації завдяки своїй міцності і біофункціональній сумісності. Його переваги стають більш очевидними зі збільшенням продуктивності та зниженням вартості технології синтезу ДНК і секвенування. У

порівнянні зі зберіганням ДНК *in vitro*, зберігання *in vivo* використовує переваги ефективних клітинних механізмів реплікації ДНК, коректура і підтримки довголанцюгової ДНК дає можливість довільного доступу до даних і підтримує запис біохімічних даних в реальному часі подій *in situ* в живих організмах [5]. Методи молекулярного запису та зберігання даних у ДНК *in vivo* наведено на рис.4 [25].

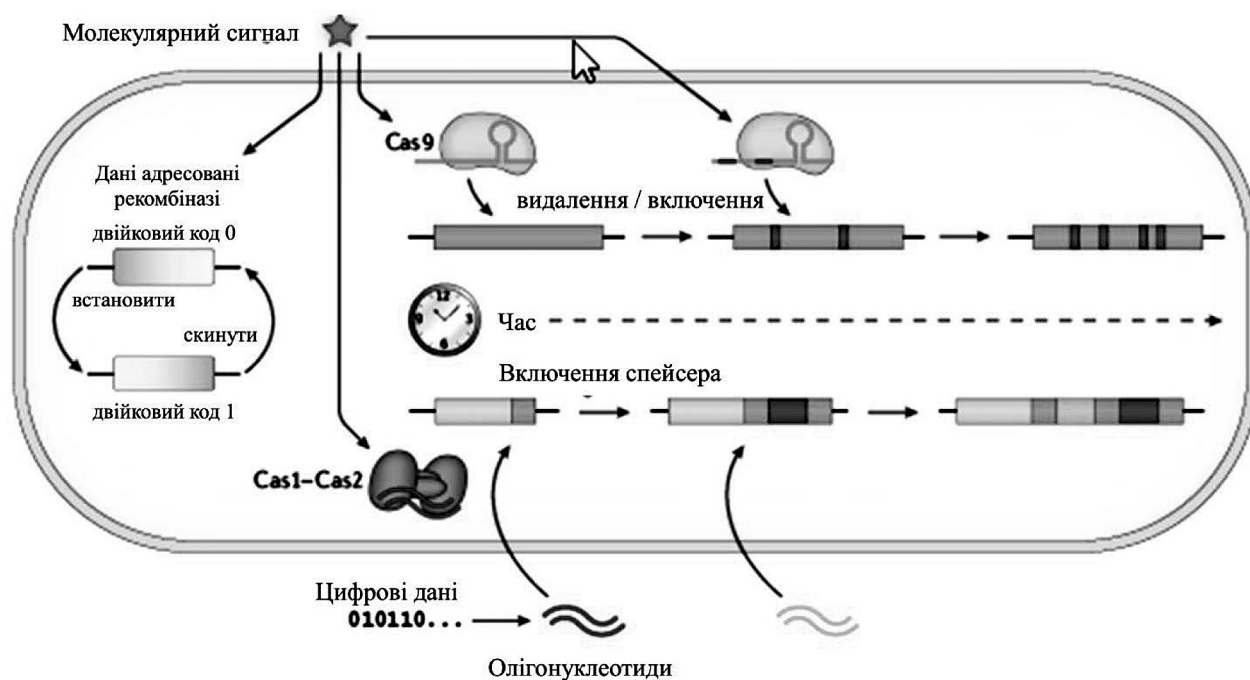


Рис. 4 . Методи молекулярного запису та зберігання даних у ДНК *in vivo*[25]

Методи запису, засновані на CRISPR (clustered regularly interspaced short palindromic repeats - кластерні регулярні взаємопромійні короткі паліндромні повтори - термін, який є синонімом редагування генома) – Cas9, використовують Cas9 та CRISPR для націлювання РНК на певні ділянки в геномі для редагування. Розщеплення цих ділянок Cas9 призводить до невеликих вставок або делецій, які можна використовувати для реєстрації величини та тривалості молекулярних сигналів з часом. У методах CRISPR – Cas1 – Cas2 інтеграційний комплекс Cas1 – Cas2 вставляє короткі шматочки ДНК (приблизно 30 п.о.), що називаються прото-прокладками, у певний локус, відомий як масив CRISPR.

Створення технології редагування генів CRISPR дало нові можливості розробки нових систем запису даних. З її допомогою інформацію можна записувати, зберігати та видавати у вигляді масиву даних в геномі бактерії. Запропонований

спосіб зберігання інформації в ДНК заснований на здатності записувати інформацію в геном живої клітини шляхом послідовного додавання нуклеотидів. CRISPR дозволяє хронологічно реєструвати цифрову інформацію (зокрема, зображення нерухомих і рухомих об'єктів) в живих бактеріях. CRISPR допомагає бактеріям розвивати імунітет проти постійного натиску вірусів. Використовуючи інтеграли Cas1 - Cas2, мікробна імунна система CRISPR - Cas зберігає нуклеотидну інформацію вірусів, які вторглися, для забезпечення адаптивного імунітету. Ця система може записувати необхідну інформацію в геном. Для зберігання інформації про інфекції вона захоплює молекули вірусної ДНК і генерує з них короткі послідовності, які поступово накопичуються. Таким чином виходить, що постійно зростаючий масив інформації - ця «антивірусна бібліотека», розташована в локусі CRISPR в геномі бактерії. Широко застосований сьогодні білок CRISPR-Cas9 використовує цю пам'ять для

знищення тих же вірусів, коли вони повертаються. Однак інші частини системи CRISPR до сих пір не привертали особливої уваги (мабуть, крім Cas13a - цю нуклеазу використовували для створення тесту SHERLOCK, який "полює" на патогени і мутації). Для запису інформації код ДНК, який з додатковими послідовностями включається в спейсери (ділянки ДНК, що не транскриптуються та розташовані між генами, які тандемно повторюються). Потім дані розміщують у хронологічному порядку в масивах CRISPR в геномах кишкової палички. Використання CRISPR - Cas може бути застосовано для кодування значень пікселів чорно-білих зображень і короткометражного фільму в геномі популяції живих бактерій. Була використана технологія CRISPR / Cas9, яка дозволяє вставити в геном кишкової палички нуклеотиди, відповідні пікселям зображення. Ця система може збирати і стабільно зберігати практичні обсяги реальних даних в геномах популяцій живих клітин [23-25]. Спочатку в ДНК бактерії було вбудовано зображення людської руки і була продемонстрована можливість запису статичного зображення (рис.5). Також була записана класична анімація, створена Едвардом Мейбрідж в 1878 р., що складається з п'яти кадрів, що зображують, коня що скаче [23]. Результати записи зображень в ДНК пам'яті на основі технології CRISPR представлені на рис. 5. [23].

Особливу увагу дослідники звертають на те, що їм вдалося відновити кадри анімації рівно в тому порядку, в якому вони були записані, оскільки бактерії зберігають нові записи чіткіше попередніх. Це визначає що закладена в ДНК мікроорганізмів інформація може бути відтворена в точній послідовності від початку до кінця, що робить бактерії практично ідеальним сховищем будь-якої інформації, навіть відеороликів, для яких правильна послідовність кадрів критично важлива [24-25].

Використання технології CRISPR для зберігання даних має ряд недолік, зокрема, існує ймовірність мутацій в мікроорганізмах. Як відомо, клітини - це живі істоти, які мають властивість гинути. Звичайно, так як інформація, що зберігається, розташована в геномі, бактерії зможуть передавати її у спадок, проте кожен акт розмноження обумовлений певними ризиками, в тому числі - ризиком мутації ДНК, яка може просто знищити дані, що зберігаються. Якщо використовувати для зберігання інформації велику популяцію клітин, то можливі мутації - ймовірність яких не так велика, як може здатися на перший погляд - в цілому не повинні

суттєво вплинути на закодовану інформацію. Крім цього, на даний момент остаточно не зрозуміло, як багато інформації може вмістити в собі ДНК живої бактерії, особливо в порівнянні з синтетичною ДНК. Технологія CRISPR для зберігання даних суттєво перевершує зберігання даних у синтетичній ДНК в тому відношенні, що дозволяє не тільки зберігати інформацію, але й змінювати або доповнювати її з часом [23].

Запропоновано оригінальну технологію запису, при якій ділянку синтетичної молекули ДНК з цифровим кодом збирають по літерах, а потім за допомогою бактерій впроваджують в рідну ДНК рослини. В процесі експерименту вчені закодували повідомлення «Привіт, світ!» в насіння тютюну. Нові рослини, вирощені з насіння, містили модифіковану ДНК в кожній клітині. Для того щоб відтворити інформацію, вчені витягували ДНК з рослин і секвенували її за допомогою традиційних методів генетичного аналізу. Після того як дані перевели назад в двійкову систему, на екрані з'явилося повідомлення: «Привіт, світ!». Технологія стане основою «інтелектуальних» дерев і рослин, які будуть сховищами багатоміжових знань. Встановлено, що на одному ящику насіння можна зберегти всі існуючі зараз архіви [26].

Методи кодування даних в ДНК пам'яті. Складність записи даних в ДНК пам'яті полягає в тому, щоб перевести велику кількість байтів - стандартних одиниць даних, спеціально розроблених для лінійних і послідовних сховищ, таких як оперативна пам'ять і жорсткі диски, в закручені нанорозмірні структури ДНК. Переклад одного формату даних в інший виявляється досить складним [27].

Класичний підхід до запису даних в ДНК передбачає перетворення послідовності бітів - нулів і одиниць - в послідовність з чотирьох базових основ ДНК. Наприклад, азотисті основи А, Т, G, С можна представити так: А = 00, Т = 01, G = 10, С = 11. Користуючись цим підходом, в 2016 р. компанії Microsoft вдалося записати 200 Мбайт тексту і відео в синтетичних молекулах ДНК [12]. Однак такий спосіб погано підходить для масової записи даних внаслідок високої ціни [7]. Church G. et al використовували одну підставу для подання єдиною двійковою цифрою (тобто А, С = 0; G, Т = 1) [11]. Однак низька щільність інформації завадила його використанню в більш пізніх дослідженнях. Goldman N. et al. запропонували спосіб, який полягає в тому, щоб відмовитися від четвіркової системи (Base-4) на користь трійкової

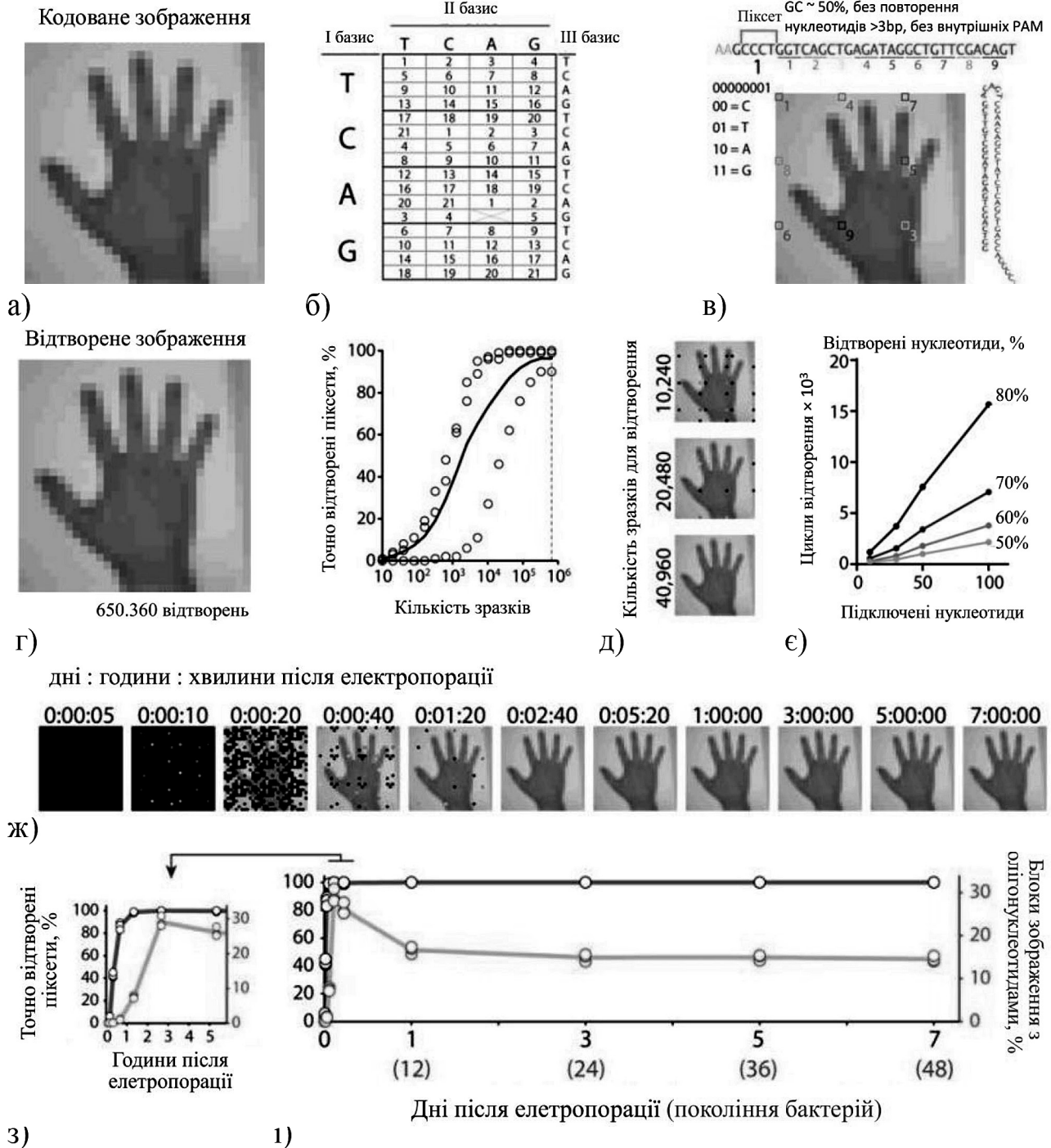


Рис. 5. Зображення записане у геномі: а) зображення руки; б) кодування 21 кольору; в) послідовність вгорі показує лінійний протоспейсер з кодом пікселя, за яким слідують значення пікселів (розподілені по зображенню). Піксель показаний під нуклеотидами з перетворенням двійкового коду в нуклеотид. Маленькі кольорові числа під протоспейсером позначають окремі пікселі на зображенні. Мінімальний протоспейсер шпильки показаний праворуч; г) одна репліка при 655360 читаннях. Чорний відображається, якщо інформація про пікселі не відновлені; д) точно відтворені пікселі по глибині читання. Незафарбовані гуртки показують точки з окремих повторів, чорна лінія показує середнє значення; е) результат впливу дискретизації секвенування; ж) кількість відтворень, що необхідні для досягнення точності 50, 60, 70 і 80% для даного набору олігонуклеотидів; з) послідовності зображень в моменти часу після електропорації; і) кількісна оцінка відсотка точно відкликаних пікселів (чорним кольором) і відсотка масивів з оліго-похідними спейсерами (червоним кольором) по часовим точкам. Незафарбовані кружки представляють окремі повтори, лінії показують середнє значення. Вирізка графіка (зліва) розширює перші шість годин [23]

(Base-3), а четвертий нуклеотид використовувати в службових цілях для розбиття довгих ланцюжків (СССАСССАСССАСССАССС). Цей підхід повністю уникає гомополінуклеотидних (множинних копій одного і того ж нуклеотиду) записів без шкоди для щільності інформації [11].

Більшу щільність запису забезпечує завадостійке фонтанне кодування (Digital Fountain Codes), запропоноване для використання у ДНК пам'яті [22]. Кодами з цього класу можна закодувати повідомлення

кінцевого розміру потенційно-необмеженим потоком незалежних пакетів. Незалежність генерування кодів символів забезпечується застосуванням статистичного кодування. Відмінна властивість фонтанних кодів полягає в тому, що їх застосування дозволяє відправляти дані по ненадійним каналам зв'язку [5,22]. В випадку фонтанного кодування ДНК вихідний блок даних поділяється на кілька частин різної довжини - інформаційні символи. Схему фонтанного кодування у ДНК пам'яті наведено на рис.6

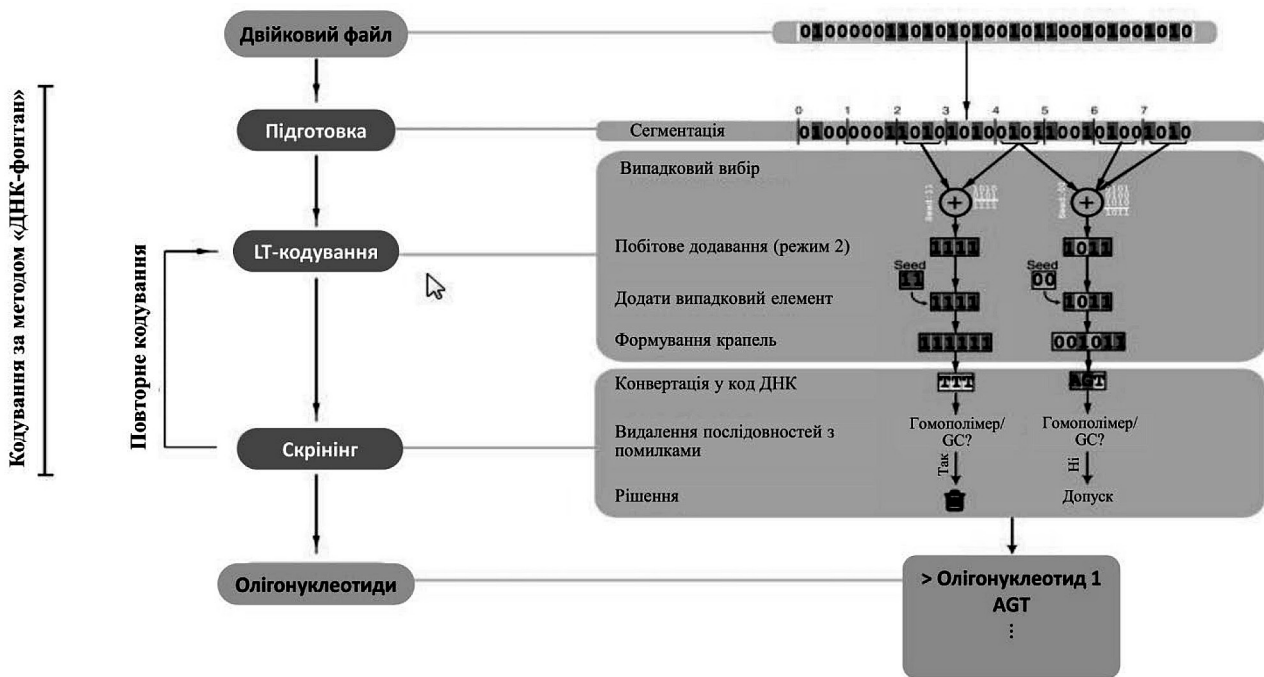


Рис. 6. Ліворуч – три основні кроки алгоритму. Праворуч – приклад фонтанного кодування малого файлу об'ємом 32 біта. Для простоти файл розділений на 8 сегментів по 4 біта [28]

Кодер працює у три етапи (рис. 6). По-перше, він попередньо обробляє двійковий файл у ряд сегментів певної довжини, що не перекриваються. Далі виконується двома етапа обчислень: перетворення Luby (Luby transform) та скрінінг. Перетворення Luby створює основу фонтанних кодів. В основному він формує дані у вигляді бажаної кількості коротких повідомлень, які називаються краплями, шляхом виділення випадкової підмножини сегментів із файлу за допомогою спеціального розподілу та додавання їх побітовим чином у двійковому полі. Крапля містить два інформаційних блоки: один з блоків містить корисний набір даних, що містить результат процедури додавання, та коротке початкове число фіксованої довжини. Це коротке початкове число відповідає стану генератора випадкових чисел перетворення під час створення

краплі і дозволяє декодеру визначити ідентичність сегментів у краплі. У фонтанному кодуванні ДНК алгоритм застосовує один раунд перетворення в кожній ітерації, щоб створити одну краплю. Далі алгоритм переходить до стадії скрінінгу сортування крапель. Цей етап дозволяє повністю реалізувати потенціал кодування кожного нуклеотиду. Під час скрінінгу алгоритм перетворює двійкову краплю в послідовність ДНК, перетворюючи {00,01,10,11} в {A, C, G, T}, відповідно. Потім він перевіряє послідовність бажаних біохімічних властивостей вмісту GC і гомополімерів. Якщо послідовність проходить екран, вона вважається дійсною та додається до файлу дизайну олігомеру; в іншому випадку алгоритм просто не враховує крапельку. Оскільки перетворення Luby може створити будь-яку бажану кількість крапель,

процес створення та відбору крапель продовжується, поки не буде сформовано достатню кількість дійсних нуклеотидів. На практиці рекомендується використовувати на 5-10% більше нуклеотидів ніж вхідних сегментів. Для аналізу використання фонтанного кодування у ДНК пам'яті було проаналізовано обмеження на вміст GC в молекулах ДНК, оскільки вміст GC впливає на стабільність молекул ДНК, частоту виникнення помилок заміщення і на час секвенування. Послідовності ДНК з високим вмістом GC або тривалими гомополімерними циклами (наприклад, АААААА ..) є небажаними, оскільки їх важко синтезувати і призводить до помилок секвенування. Вміст GC, близький до 50%, може бути отримано шляхом вибору молекул ДНК з відповідними співвідношеннями GC для зберігання інформації, відкидаючи молекули з невідповідним вмістом GC. Кількісний аналіз показує, що біохімічні обмеження знижують кодувальний потенціал кожного нуклеотиду до 1,98біт / на нуклеотид для максимальної здатності кодування у ДНК пам'яті [22,28].

В даний час проводяться дослідження з розробки методів кодування, які можуть бути використані для масового запису даних, при цьому їх вартість буде невисокою. Замість того, щоб використовувати мільйони ДНК-ланцюжків, пропонується генерувати велику кількість різних ДНК-молекул, що складаються не більше ніж з 30п.о.(пар основ). Потім за рахунок ферментативних реакцій ці попередньо підготовлені «шматочки» формують особливі патерни, які і кодують інформацію. Таким чином, замість того щоб бути однією з азотистих основ, біти розміщуються в багатовимірні матриці, а групи молекул відображають становище бітів в цих матрицях [7].

Для того, щоб зробити працездатний стандарт зберігання даних на ДНК потрібно вирішити завдання виявлення помилок зберігання даних – заміни, вставки та вилучення. Помилки першого типу подібні зміні значення бітів - це рівнозначно тому, що 0 стає 1, або 1 стає 0, і це легко виявляється та виправляється за допомогою коду Ріда-Соломона. Проблема полягає в тому, що в ДНК чотири нуклеотиди – А, Г, Ц, Т, і кожен з них може помилково записатися будь-яким іншим, що серйозно збільшує кількість можливих помилок і ускладнює можливість їх виявлення та виправлення. Два інших класи помилок представляють собою випадки, коли пари основ ДНК додаються або видаляються з ланцюга. Принципово важливе

те, що для ДНК немає надійного природнього способу контролю наявності у ланцюгу нуклеотидів помилок заміщення, вставки чи вилучення. Однозначний характер зберігання даних у ДНК, по суті, є ключовим для протоколу HEDGES (Hash Encoded, Decoded by Greedy Exhaustive Search). Жоден ізольований нуклеотид не містить корисних даних. Група послідовних нуклеотидів забезпечують надійну систему зберігання даних, яка, дозволила досягти високої щільності зберігання даних на ДНК, що зберігаються протягом тривалого часу.

Протокол HEDGES розділяє вхідну інформацію на тисячі або мільйони коротких послідовностей по сто нуклеотидів, кожна з яких містить дані, необхідні для повторного «збирання» вихідного тексту - навіть з невідомою кількістю помилок заміни, вставки і вилучення. Розшифровування даних з їх сховища ДНК передбачає спочатку секвенування геному, а потім перетворення цих генетичних даних в біти. Після цього залишалось лише з'ясувати, які біти є «адресними», і з їх допомогою зв'язати інформаційні біти в єдиний файл даних. Автори проекту вважають, що однією з найпривабливіших особливостей їх нового протоколу є його стійкість до технологічних змін і змін формату даних протягом століть [29].

Для врахування специфічних вимог до зберігання на ДНК вважається за необхідне адаптувати існуючі схеми кодування даних: індивідуальні послідовності індексуються та використовуються два незалежні коди для виправлення помилок (зокрема коди Ріда-Соломона) [30].

Строки зберігання даних в ДНК пам'яті. ДНК, що знайдені в кістках вимерлих істот, являють собою емпіричні докази стабільності ДНК протягом тисяч років [31]. У печерах Нової Зеландії ідентифікували ДНК 29-ти таксонів давніх рослин та двох видів птаха моа, а у вічній мерзлоті Сибіру – ДНК мамонта, бізона, дикого коня та 19-ти таксонів рослин. В сухій печері на південно-заході США виявили ДНК вимерлого гігантського лінивця, кондора та інших тварин плейстоцену [32]. Молекули ДНК природним чином розпадаються з характерним періодом напіврозпаду, що призводить до поступової втрати інформації, яка зберігається [5,31,33,34]. Існує декілька типів пошкодження ДНК, що відбуваються в біологічних зразках як через вплив навколишнього середовища, так і через процеси у клітинах. Ці типи пошкоджень різняться залежно від фізіологічних умов та умов зберігання зразків і включають модифікацію основи, неправильні

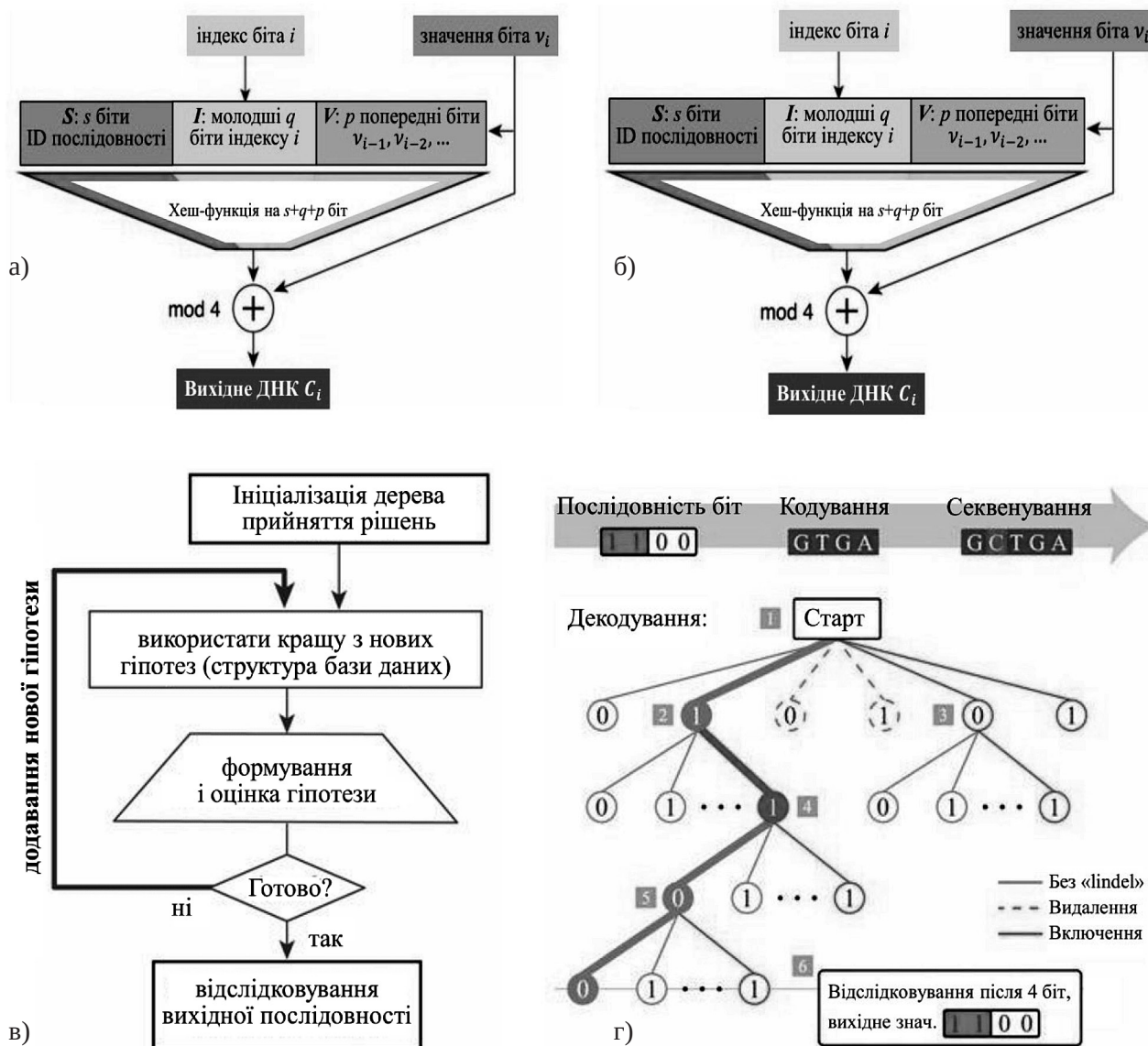


Рис.7. Принцип кодування і декодування по протоколу HEDGES [29]

пари, зшиті нуклеотиди та дволанцюгові розриви ДНК. Фрагментація довгих ділянок ДНК на більш короткі становить більшість змін властивостей, які відбуваються під час деградації ДНК [5,30]. Період напіврозпаду ДНК корелює з температурою і довжиною фрагменту ДНК. Оскільки синтез і секвенування дуже довгих ланцюгів ДНК технічно більш складні, дані доцільно зберігатися на кількох коротких сегментах ДНК [35]. Було показано, що молекула ДНК розміром 500 бр має період напіврозпаду 30 р. при 25 °С та до 500 р. для фрагмента 30 бр [33]. ДНК у кістках набагато краще зберігаються, якщо вони інкапсульовані у кристалічні структури типу апатиту, які захищають тверду ДНК від впливу навколишнього середови-

ща [30]. Збереженню ДНК сприяє з'єднання її з оточуючими сполуками, наприклад із глинистими мінералами, великими органічними молекулами, які захищають адсорбовану ДНК від дії нуклеаз [32]. Загальний підхід у стабілізації характеристик та зберігання біологічних зразків при кімнатній температурі полягає у створенні термостабільного бар'єру навколо біологічного зразка, щоб захистити його від деградації [35]. Часова стабільність ДНК значно поліпшується за рахунок зберігання при низьких температур і забезпеченні захисту від вологи [31]. Так, при -5 °С період напіврозпаду фрагмента мітохондріальної ДНК довжиною 30 бр в кістці, за прогнозами, складе 158000 років. [31]. Передбачається, що пам'ять на ДНК можна

розмістити в надзвичайно холодних регіонах Землі для зберігання інформації протягом тисячоліть [5].

Для забезпечення стабільності часових характеристик ДНК пам'яті пропонується здійснювати герметизацію молекул ДНК [5,30,36,37]. Кращі

характеристики за строками зберігання даних має пам'ять на твердотільній ДНК або ДНК в рідких середовищах [30]. Схема процесу зберігання даних на ДНК з герметизацією молекул ДНК діоксидом кремнію представлена на рис.8. [30].

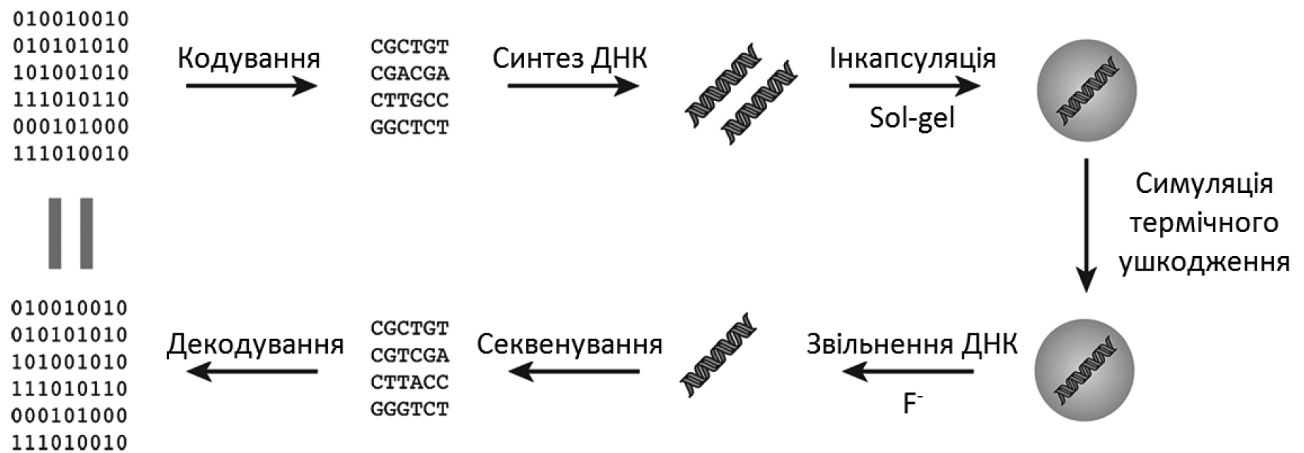


Рис.8. Схема запису та зберігання інформації на герметизованій ДНК [30]

Для зчитування даних діоксид кремнію розчиняється сполуками, що містять фтор, і ДНК зчитується за допомогою секвенування. У дослідженні Allentoft M. [30] секвенування здійснювалось на обладнанні компанії Illumina- (<https://www.illumina.com>). Відтворення даних здійснювалось навіть за умов появи помилок під час процедур запису даних [30]. Термін зберігання записаних даних на герметизованій ДНК може становити 2000 років при $9,4^{\circ}\text{C}$ або 2 мільйони років при -18°C , що перевершує показники інших носіїв для довгострокового зберігання даних [30]. Тривалий термін зберігання даних в ДНК, навіть при кімнатній температурі, робить зберігання інформації ДНК особливо ефективним для «холодних» даних з доступом, який відбувається рідко [5].

Висновки. 1. Розроблені методи зберігання даних на ДНК дозволяють організувати поки що лише постійне зберігання даних і тому придатні тільки для зберігання інформації, яка не потребує модифікації, зокрема архівної інформації.

2. В першу чергу зберігання даних на ДНК знайде місце при реалізації унікальних особливостей систем запису, таких як стеганографічний запис інформації тощо.

3. Масштабна реалізація зберігання даних на ДНК може бути здійснена при збільшенні швидкості запису і відтворення даних та зменшенні вартості зберігання даних. Головним шляхом вдосконалення характеристик ДНК пам'яті є створення більш ефективних систем кодування даних.

Література.

1. Long-term storage of digital information / V. V. Petrov, Zichun Le., A. A. Kryuchyn, S. M. Shanoylo et al. — P. 148. — doi: 10.15407/Akademperiodyka. 360.148.
2. Dynamic and scalable DNA-based information storage / K. N. Lin, K. Volkel, J. M. Tuck, A. J. Keung // Nature Communications. — 2020. — Vol. 11. — Article number: 2981. — Режим доступу: <https://doi.org/10.1038/s41467-020-16797-2>.
3. Способи вирішення проблеми довгострокового зберігання інформації, записаної в цифровому вигляді / В. В. Петров, А. А. Крючин, С. М. Шанойло, І. О. Косско, В. Г. Кравець // Доповіді Національної академії наук України. — 2003. — № 4. — С. 52-58.
4. Стан і проблеми створення ДНК-пам'яті / А. А. Крючин, Є. В. Беляк, Є. А. Крючина, А. В. Потєбня // Медична інформатика та інженерія. — 2015. — № 3 (31). — С. 9-16.
5. DNA storage: research landscape and future prospects / Y. Dong, F. Sun, Z. Ping et al. // National Science Review. — 2020. — Vol. 7, Issue 6. — P. 1092-1107.
6. Potomac Institute for Policy Studies // The Future of DNA Data Storage. [Електронний ресурс]. — Режим доступу: https://potomacinstitute.org/images/studies/Futureof_DNAData_Storage.pdf.
7. The Rise of DNA Data Storage Could DNA as an archival medium be the solution to our information overload? [Електронний ресурс]. — Режим доступу: <https://www.wired.com/story/the-rise-of-dna-data-storage/>.
8. Trends to store digital data in DNA: an overview / F. Akram, I. Ul. Haq, H. Ali, A. T. Laghari // Molecular Biology Reports. — 2018. — Vol. 45. — P. 1479-1490.
9. New Trends of Digital Data Storage in DNA / P. Y. De Silva, G. U. Ganegoda // BioMed research international. — 2016. — Vol. 20. — Режим доступу: <https://doi.org/10.1155/2016/8072463>.
10. Probing the physical limits of reliable DNA data retrieval / L. Organick, Y. J. Chen, S. D. Ang et al. // Nature communications. — 2020. — Vol. 11 (1). — Режим доступу: <https://doi.org/10.1038/s41467-020-14319-8>.
11. Next-Generation Digital Information Storage in DNA / G. M. Church, Y. Gao, S. Kosuri // Science. — 2012. — Vol. 337, Issue 6102. — P. 1628.
12. Langston J. With a “hello,” Microsoft and UW demonstrate first fully automated DNA data storage / J. Langston. — March 21, 2019. [Електронний ресурс]. — Режим доступу: <https://news.microsoft.com/innovation-stories/hello-data-dna-storage/>.
13. Shankland S. Startup packs all 16GB of Wikipedia onto DNA strands to demonstrate new storage tech — Biological molecules will last a lot longer than the latest computer storage technology, Catalog believes / S. Shankland // CNET. Retrieved, 7 August 2019.
14. Random access in large-scale DNA data storage / L. Organick, S. Ang, Y. J. Chen et al. // Nat Biotechnol. — 2018. — Vol. 36. — P. 242-248.
15. DNA-of-things storage architecture to create materials with embedded memory / J. Koch, S. Gantenbein, K. Masania et al. // Nat Biotechnol. — 2020. — № 38. — P. 39-43.
16. «DNA of Things» — Storing Extensive Data in Everyday Objects [Електронний ресурс]. — Режим доступу: <https://scitechdaily.com/dna-of-things-storing-extensive-data-in-everyday-objects/>.
17. Rapid authentication of pharmaceuticals via DNA tagging and field detection / L. Jung, M. E. Hogan, Y. Sun et al. // PLoS ONE. — 2019. — Vol. 14 (6). — P. e0218314.
18. Quantifying molecular bias in DNA data storage / Y. Chen, C. N. Takahashi, L. Organick et al. // Nature Communications. — 2020. — Vol. 11. — Режим доступу: <https://doi.org/10.1038/s41467-020-16958-3>.
19. Random access DNA memory in a scalable, archival file storage system / J. L. Banal, T. R. Shepherd, J. Berleant et al. — Режим доступу: <https://www.biorxiv.org/content/10.1101/2020.02.05.936369v2.full>.
20. Forward Error Correction for DNA Data Storage / M. Blawat, K. Gaedke, I. Hütter et al. // Computer Science. — 2016. — Vol. 80. — P. 1011-1022.
21. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA / N. Goldman, P. Bertone, S. Chen et al. // Nature. — 2013. — Vol. 494. — P. 77-80.
22. DNA Fountain enables a robust and efficient storage architecture / Y. Erlich, D. Zielinski // Science. — 2017. — Vol. 355 (6328). — P. 950-954.
23. CRISPR — Casencoding of a digital movie into the genomes of a population of living bacteria / S. Shipman, J. Nivala, J. Macklis et al. // Nature. — 2017. — Vol. 547. — P. 345-349.
24. CRISPR — cas system: biological function in microbes and its use to treat antimicrobial resistant pathogens / Shabbir M. Abu Bakr, Q. Wu, S. Mahmood et al. // Annals of Clinical Microbiology and Antimicrobials. — 2019. — Vol. 18. — Режим доступу: <https://doi.org/10.1186/s12941-019-0317-x>.
25. Molecular digital data storage using DNA / L. Ceze, J. Nivala, K. Strauss // Nature Reviews Genetics [Електронний ресурс]. — Режим доступу: <https://www.gwern.net/docs/genetics/editing/2019-ceze.pdf>.
26. The Potential of Plants and Seeds in DNA-Based Information Storage / K. Fister, Jr. I. Fister, J. Murovec // In book: Understanding Information, 2017. — P. 69-81.
27. DNA Memory / Arita M., Hagiya M., Takinoue M., Tanaka F. — Handbook of Natural Computing. — Springer, 2012.
28. Capacity — approaching DNA storage / Y. Erlich, D. Zielinski. — 2016. — Режим доступу: <https://doi.org/10.1101/074237>.
29. HEDGES error-correcting code for DNA storage corrects in dels and allows sequence constraints / W. H. Press, J. A. Hawkins, S. K. Jones Jr et al. // PNAS. — 2020. — Vol. 117 (31). — P. 18489-18496.

30. Robust Chemical Preservation of Digital Information on DNA in Silica with Error — Correcting Codes / R. N. Grass, R. Heckel, M. D. Paunescu et al. // *Angew Chem Int Ed.* — 2015. — Vol. 54. — P. 2552-2555.
31. Protocols for dry DNA storage and shipment at room temperature / N. V. Ivanova, M. L. Kuzmina // *Mol Ecol Resour.* — 2013. — Vol. 13 (5). — P. 890-898.
32. ДНК оточуючого середовища: екологічний та генетичний аспекти / В. М. Помогайбо, Л. Д. Орлова, Н. О. Власенко // *Ecology and noospherology.* — 2016. — Vol. 27, no. 1-2. — P. 16-24.
33. The half-life of DNA in bone: measuring decay kinetics / M. E. Allentoft, M. Collins, D. Harker et al. // *Proc. R. Soc. B.* — 2012. — Vol. 279. — P. 4724-4733.
34. Evaluation of DNA stable™ for DNA storage at ambient temperature / S. E. Howlett, H. S. Castillo, L. J. Gioeni et al. // *Forensic Sci Int Genet.* — 2014. — Vol. 8, Issue 1. — P. 170-178.
35. Hofreiter Molecular caving / M. Hofreiter, J. I. Mead, P. Martin, H. N. Poinar // *Curr. Biol.* — 2003. — Vol. 13 (18). — P. 693-695.
36. Emerging applications for DNA writers and molecular recorders / F. Farzadfard, T. K. Lu // *Science.* — 2018. — Vol. 361 (6405). — P. 870-875.
37. Silica Microcapsules for Long-Term, Robust, and Reliable Room Temperature RNA Preservation / M. Puddu, W. J. Stark, R. N. Grass // *Adv Healthc Mater.* — 2015. — Vol. 4 (9). — P. 1332-1338.
9. De Silva P. Y., Ganegoda G. U. (2016). New Trends of Digital Data Storage in DNA. *BioMed research international*, Vol. 2016. <https://doi.org/10.1155/2016/8072463>.
10. Organick L., Chen Y. J., Dumas Ang S., Lopez R., Liu X., Strauss K., Ceze L. (2020). Probing the physical limits of reliable DNA data retrieval. *Nature communications*, 11(1). <https://doi.org/10.1038/s41467-020-14319-8>.
11. Church G. M., Gao Y., Kosuri S. (2012). *Next-Generation Digital Information Storage in DNA. Science*, 337:6102, 1628.
12. Langston J. (2019). With a «hello,» Microsoft and UW demonstrate first fully automated DNA data storage. <https://news.microsoft.com/innovation-stories/hello-data-dna-storage/>.
13. Shankland S. (2019). Startup packs all 16GB of Wikipedia onto DNA strands to demonstrate new storage tech - Biological molecules will last a lot longer than the latest computer storage technology, Catalog believes. CNET.
14. Organick L., Ang S., Chen Y. J. et al. (2018). Random access in large-scale DNA data storage. *Nat Biotechnol.*, 36, 242-248.
15. Koch J., Gantenbein S., Masania K. et al. (2020). DNA-of-things storage architecture to create materials with embedded memory. *Nat Biotechnol*, 38, 39-43. <https://doi.org/10.1038/s41587-019-0356-z>.
16. «DNA of Things» – Storing Extensive Data in Everyday Objects. <https://scitechdaily.com/dna-of-things-storing-extensive-data-in-everyday-objects/>.
17. Jung L., Hogan M. E., Sun Y. et al. (2019). Rapid authentication of pharmaceuticals via DNA tagging and field detection. *PLoS ONE*, 14(6), e0218314.
18. Chen Y., Takahashi C. N., Organick L. et al. (2020). Quantifying molecular bias in DNA data storage. *Nature Communications*, 11.
19. Banal J. L., Shepherd T. R., Berleant J. et al. (2020). Random access DNA memory in a scalable. doi: 10.1101/2020.02.05.936369 PPR: PPR112067.
20. Blawat M., Gaedke K., Hütter I. (2016). Forward Error Correction for DNA Data Storage. *Procedia Computer Science*, 80, 1011-1022.
21. Goldman N., Bertone P., Chen S. et al. (2013). Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, 494, 77-80. doi:10.1038/nature11875.
22. Erlich Y., Zielinski D. (2017). DNA Fountain enables a robust and efficient storage architecture. *Science*, 355 (6328), 950-954. doi: 10.1126/science.aaj2038. PMID: 28254941.
23. Shipman S., Nivala J., Macklis J. et al. (2017). CRISPR – Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature*, 547, 345-349.
24. Shabbir M. Abu Bakr, Wu Q., Mahmood S. et al. (2019). CRISPR – cas system: biological function in microbes and its use to treat antimicrobial resistant pathogens. *Annals of Clinical Microbiology and Antimicrobials*, 18.

References.

1. Petrov V. V., Le Zichun, Kryuchyn A. A. et al. (2018). Long-term storage of digital information. doi: 10.15407/Akadempriodyka.360.148.
2. Lin K. N., Volkel K., Tuck J. M., Keung A. J. (2020). Dynamic and scalable DNA-based information storage. *Nature Communications*. Vol. 11.
3. Petrov V. V., Kryuchyn A. A., Shanoilo S. M. et al. (2003). Ways to solve the problem of long-term storage of information recorded in digital form. *Reports of the National Academy of Sciences of Ukraine*, 4, 52-58. [In Ukrainian].
4. Kryuchyn A. A., Belyak Ie. V., Kryuchyna E. A., Potebnya A. V. (2015). The state and problems of creating DNA memory. *Medical Informatics and Engineering*, 3 (31), 9-16. [In Ukrainian].
5. Dong Y., Sun F., Ping Z., Ouyang Q., Qian L. (2020). DNA storage: research landscape and future prospects. *National Science Review*, Vol. 7, Issue 6, 1092-1107.
6. Potomac Institute for Policy Studies. *The Future of DNA Data Storage*. https://potomacinstitute.org/images/studies/Future_of_DNA_Data_Storage.pdf.
7. The Rise of DNA Data Storage Could DNA as an archival medium be the solution to our information overload?. <https://www.wired.com/story/the-rise-of-dna-data-storage/>
8. Akram F., Haq I., Ali H., Laghari A. T. (2018). *Trends to store digital data in DNA: an overview*. *Molecular Biology Reports*, Vol. 45, 1479-1490.

25. Ceze L., Nivala J., Strauss K. (2019). Molecular digital data storage using DNA. *Nature Reviews Genetics*. <https://www.gwern.net/docs/genetics/editing/2019-ceze.pdf>.
26. Fister K., Fister Jr. I., Murovec J. (2017). The Potential of Plants and Seeds in DNA – Based Information Storage. In book: *Understanding Information*, 69-81. doi: 10.1007/978-3-319-59090-5_4.
27. Arita M., Hagiya M., Takinoue M., Tanaka F. (2012). DNA Memory. In: Rozenberg G., Bäck T., Kok J. N. (eds). *Handbook of Natural Computing*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-92910-9_38.
28. Erlich Y., Zielinski D. Capacity-approaching DNA storage. (2016). doi: <https://doi.org/10.1101/074237>.
29. Press W. H., Hawkins J. A. et al. (2020). HEDGES error-correcting code for DNA storage corrects indels and allows sequence constraints. *PNAS*, 117 (31), 18489-18496.
30. Grass R. N., Heckel R., Paunescu M. D. et al. (2015). Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes. *Angew Chem Int Ed.*, 54, 2552-5.
31. Ivanova N.V., Kuzmina M. L. (2013). Protocols for dry DNA storage and shipment at room temperature. *Mol Ecol Resour.*, 13 (5), 890-898. doi: 10.1111/1755-0998.12134.
32. Pomogaibo V. M., Orlova L. D., Vlasenko N. O. (2016). DNA of the environment: ecological and genetic aspects. *Ecology and noospherology*, 27, 1-2, 16-24. [In Ukrainian].
33. Allentoft M. E., Collins M., Harker D. et al. (2012). The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B.*, 279, 4724-4733.
34. Howlett S. E., Castillo H. S., Gioeni L. J. (2014). Evaluation of DNA stable™ for DNA storage at ambient temperature. *Forensic Sci Int Genet*, 8, 1, 170-178.
35. Hofreiter M., Mead J. I., Martin P., Poinar H.N. (2003). Molecular cloning. *Curr. Biol.*, 13 (18), 693-695. doi: 10.1016/j.cub.2003.08.039.
36. Farzadfard F., Lu T. K. (2018). Emerging applications for DNA writers and molecular recorders. *Science*, 361 (6405), 870-875. doi: 10.1126/science.aat9249.
37. Puddu M., Stark W. J., Grass R. N. (2015). Silica Microcapsules for Long-Term, Robust, and Reliable Room Temperature RNA Preservation. *Adv Healthc Mater.*, 4 (9), 1332-1338. doi: 10.1002/adhm.201500132.