

Поляков Павел Сергеевич, студент факультета довузовской подготовки и среднего профессионального образования, ФГБОУ ВО «Мордовский государственный университет им. Н. П. Огарёва»
e-mail: paw.polyakov2010@yandex.ru

Росляева Мария Николаевна, студентка факультета довузовской подготовки и среднего профессионального образования, ФГБОУ ВО «Мордовский государственный университет им. Н. П. Огарёва»
e-mail: mar.rosliaewa@yandex.ru

Прокин Александр Александрович, преподаватель факультета довузовской подготовки и среднего профессионального образования, ФГБОУ ВО «Мордовский государственный университет им. Н. П. Огарёва»
e-mail: aaprokin90@yandex.ru

ОБЗОР ПОСЛЕДНИХ ТЕНДЕНЦИЙ В ОБЛАСТИ ЦИФРОВОГО ХРАНЕНИЯ ДАННЫХ О ДНК

Аннотация: ДНК (дезоксирибонуклеовая кислота), которая, как говорят, является образцом человеческой жизни, признана мощным средством хранения цифровой информации.

Двоичная информация цифровых данных может быть сопоставлена с четырьмя строительными нуклеотидами: аденином, гуанином, цитозином и тиминном. Очень большие данные могут храниться в пределах небольшой части ДНК. Хранение осуществляется путем кодирования цифровой информации в нуклеотиды.

Кодированная информация синтезируется для получения определенной последовательности. Полученная последовательность дополнительно декодируется для извлечения данных. Использование ДНК в качестве носителя информации имеет несколько преимуществ, таких как:

- значительное сокращение места для хранения;
- сохранение данных в течение очень длительного времени.

В то же время существует ограничение, заключающееся в том, что синтез ДНК является очень дорогим процессом. Данная научная статья рассматривает новые тенденции, принятые несколькими учеными различных институтов для хранения и извлечения данных о ДНК.

Ключевые слова: Секвенирование ДНК, синтезатор, цифровые данные.

Annotation: DNA (deoxyribonucleic acid), which is said to be a model of human life, is recognized as a powerful digital storage facility.

Binary information from digital data can be compared with four building nucleotides: adenine, guanine, cytosine and thymine. Very large data can be stored within a small fraction of DNA. This is stored by encoding the digital information into nucleotides.

The coded information is synthesized to obtain a specific sequence. The resulting sequence is further decoded for data retrieval. There are several advantages to using DNA as a storage medium, such as:

- a significant reduction in storage space;
- data retention for a very long time.

At the same time, there is the limitation that DNA synthesis is a very expensive process. This research paper looks at new trends adopted by several scientists from different institutions for storing and retrieving DNA data.

Keywords: DNA sequencing, synthesizer, digital data.

ВВЕДЕНИЕ

Ранее человечество использовало несколько традиционных носителей информации, такие как бумага, файлы и т.п. Такие традиционные носители по сей день требуют достаточно большого пространства для хранения и немалых затрат на обслуживание.

Позже человеческое общество перешло к цифровым носителям, таким как zip-диски, дискеты, компакт-диски и так далее. На дискете можно хранить 2,88 мегабайта данных, но эпоха смартфонов полностью изменила эту ситуацию, увеличив объем памяти в 44 000 раз.

Теперь можно хранить содержимое 150 смартфонов, которые имеют не менее 10 000 гигабайт данных, в чрезвычайно компактной среде хранения ДНК.

В живых организмах нуклеотиды кодируются на клетках, проецируя их поведение. Точно так же цифровая информация кодируется и наносится на коды ДНК для хранения информации. Сохраненная информация может впоследствии извлекаться посредством декодирования зашифрованных данных посредством некоторых алгоритмов.

СТРУКТУРА ДНК ПРОТИВ БИНАРНЫХ ЦИФРОВ

Комплементарные партнеры в паре оснований ДНК всегда состоят из одного пурина и одного пиримидина, а не двух пуринов или двух пиримидинов [2].

Пурины – это нуклеотиды, которые имеют структуры с двойным кольцом.

Пиримидины – это нуклеотиды, которые имеют синглерные структуры.

Простой способ различить этих «партнеров»: более длинные имена (пиримидины) имеют меньшие кольца, а более короткие имена (пурины) имеют более крупные кольца.

Все данные, поступающие в компьютерную систему, хранятся в виде двоичных значений, по-другому, в виде набора нулей и единиц. Эти двоичные значения преобразовываются в исходную форму микропроцессором. Этот механизм показывается на рисунке 1.

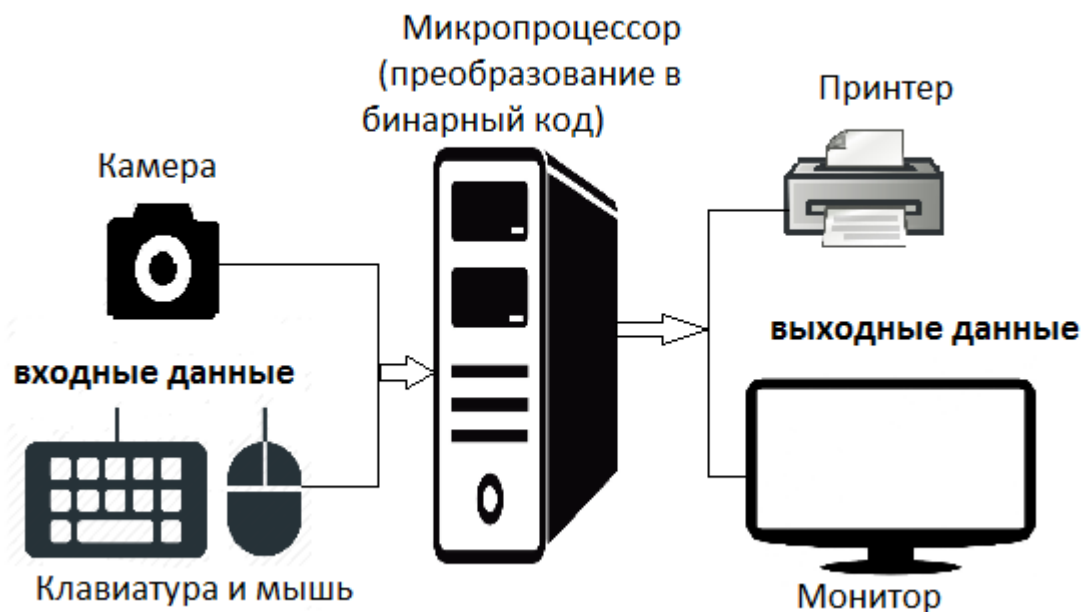


Рисунок 1 – Общий процесс преобразования данных

СИНТЕЗ ДНК

Последовательность одноцепочечной ДНК может быть синтезирована химически.

Первым этапом репликации ДНК является разрыв водородных связей между двумя анти параллельными цепями молекулы ДНК. Разделение связей между аденином и тиминном происходит достаточно легко, так как существуют только две водородные связи, тогда как разделение гуанина и цитозина затруднено из-за его трех-водородной связи. С помощью разрушающего фермента, называемого геликазой, связи разрушаются. Цепь после процесса репликации называется ветвью репликации. Происходит процесс репликации, за которым следует процесс удлинения. Вновь образованная двухцепочечная молекула ДНК содержит одну старую цепь и новую, поэтому говорят, что это полуконсервативный процесс. Когда полимераза достигает конца цепи, процесс прекращается [1].

МЕХАНИЗМ ХРАНЕНИЯ ЦИФРОВОЙ ИНФОРМАЦИИ В МОЛЕКУЛЕ ДНК

На рисунке 2 наглядно изображен механизм хранения и поиска данных в цепи ДНК. Входные данные поступают в синтезатор ДНК и хранятся в библиотеке хранения. С помощью амплификатора (ПЦР-машина) данные можно снова получить.

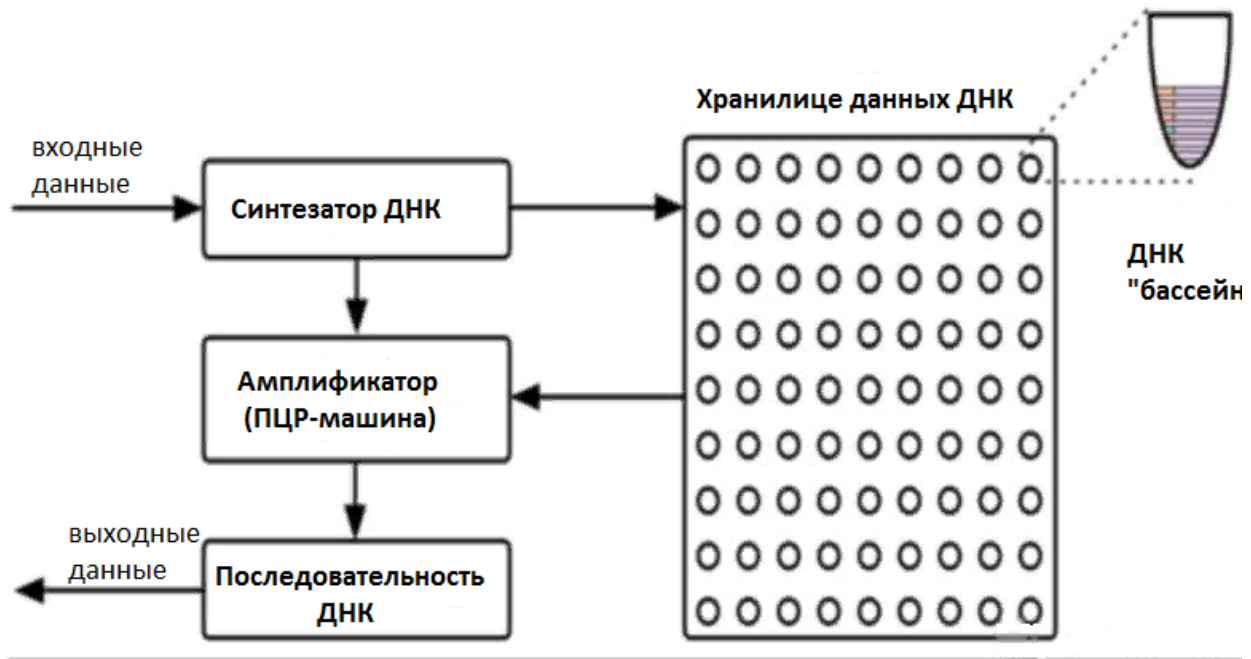


Рисунок 2 – Общий способ хранения цифровых данных на нити ДНК

СИНТЕЗАТОР ДНК

Синтезатор ДНК закодирует входные данные.

ДНК «БАССЕЙН»

«Бассейн» содержит коллекцию молекулы ДНК, которую можно сопоставить с библиотекой хранилища.

ПОСЛЕДОВАТЕЛЬНОСТЬ ДНК

Секвенсор ДНК упорядочит данные и преобразует их обратно в цифровые данные.

Кодирующий механизм

Двоичные данные преобразуются в троичные данные с помощью кодирования Хаффмана. Эти троичные данные в дальнейшем преобразуются в нуклеотидный код.

ОТЧЕТ КОМПАНИИ «MICROSOFT» 2016

Microsoft со своими исследователями и учеными сообщила о большом шаге вперед в области хранения данных ДНК. ДНК может стать лучшим способом хранения данных в долгосрочной перспективе, чем те, на которые сегодня рассчитывают компании, производящие магнитные ленты. Microsoft сохранила 200 МБ данных и продемонстрировала это 7 июля 2016 года.

ОТЧЕТ КОМПАНИИ «MICROSOFT» 2017

2 марта 2017 года, Школа инженерных и прикладных наук Колумбийского университета предложила алгоритм, предназначенный для потоковой передачи видео на сотовый телефон, который может раскрыть почти полный потенциал хранения ДНК, втиснув больше информации в четыре базовых нуклеотида, утверждают исследователи. Они демонстрируют, что эта технология также чрезвычайно надёжна.

ОТЧЕТ КОМПАНИИ «MICROSOFT» 2018

Исследователям из Microsoft в сотрудничестве с американской биотехнологической компанией Twist biotech удалось сохранить 400 мегабайт данных. Этот проект под названием «Произвольный доступ в крупномасштабном хранилище ДНК» преодолел недостаток предыдущего метода, предложенного в 2016 году. Исследователи и ученые использовали долговечную синтетическую цепь ДНК для хранения данных. Обычный метод предыдущего исследования – извлечение хранимых данных в молекуле ДНК требует секвенирования всего бассейна ДНК. Этот подход кодировал около 35 различных файлов размером 400 МБ в 13 миллионов цепочек молекул ДНК и извлекал файлы по отдельности с нулевой ошибкой.

НОВАЯ МЕТОДИКА ИССЛЕДОВАТЕЛЕЙ WATERFORD 2018

20 февраля 2018 года ирландские ученые из института Уотерфорда объявили о новой методике хранения и восстановления данных в цепи ДНК с использованием бактериальных молекул. Их опрос показывает, что у человека к 2025 году распространение данных дойдет до 160 ЗБ. Исследователи зашифровывают простое сообщение – в данном случае «Hello World» – в плазмиды и сохраняют их внутри штамма Novablue бактерий кишечной палочки E Coli., которая находится в определенном месте, которое, в последствии, и становится местом хранения архива.

Другая разновидность этой бактерии – HB101, которая является мобильной, затем разряжается и перемещается в Novablue. Как только она встречает его, плазмиды, содержащие информацию, передаются из Novablue в HB101 посредством метода ассоциации, называемого конъюгацией.

Затем HB101 плывет к инструменту, способному извлекать плазмиды и считывать информацию, которую они хранят. Движение бактерии и, следовательно, конъюгация, контролируются и создают потенциал за счет использования и размещения двух совершенно разных антибиотиков среди пространства для хранения и поиска архива.

Novablue невосприимчив к Ахромицину, тогда как HB101 способен противостоять антибиотику.

Таким образом, чтобы завершить плавание по архивному корпусу, HB101 должен сначала конъюгировать с Novablue, чтобы определить его устойчивость к ахромицину.

При этом HB101 получает информацию о плаزمидах, хранящих информацию, из Novablue. Существует целый ряд проблем, которые необходимо преодолеть, прежде чем это доказательство принципа вполне может быть использовано привычно и твердо.

ВЫВОД

В этой статье были рассмотрены различные методы хранения цифровых данных на ДНК. Последние достижения Microsoft также рассмотрены. Даже при том, что хранение ДНК имеет большие преимущества, стоимость остается главным барьером. Ученые работают над несколькими проектами по минимизации затрат на искусственный синтез ДНК.

Библиографический список:

1. Азимов А. Генетический код. От теории эволюции до расшифровки ДНК. - М., 2006.
2. С. Шривастава и Р. Бадлани, «Хранение данных в ДНК», Международный журнал по электрической энергии, вып. 2, , с. 119–124, 2014.