



Storage-D: 一个支持实用及个性化 DNA 数据存储的用户友好型平台

黄小罗^{1#*}, 崔君婷^{1#}, 强薇^{1#}, 叶建文^{3#}, 王钰¹,
谢鑫颖³, 李源真¹, 戴俊彪^{1, 2*}

¹中国科学院深圳先进技术研究院, 深圳合成生物学创新研究院, 定量合成生物学重点实验室, 广东省合成基因组学重点实验室, 深圳合成基因组学重点实验室

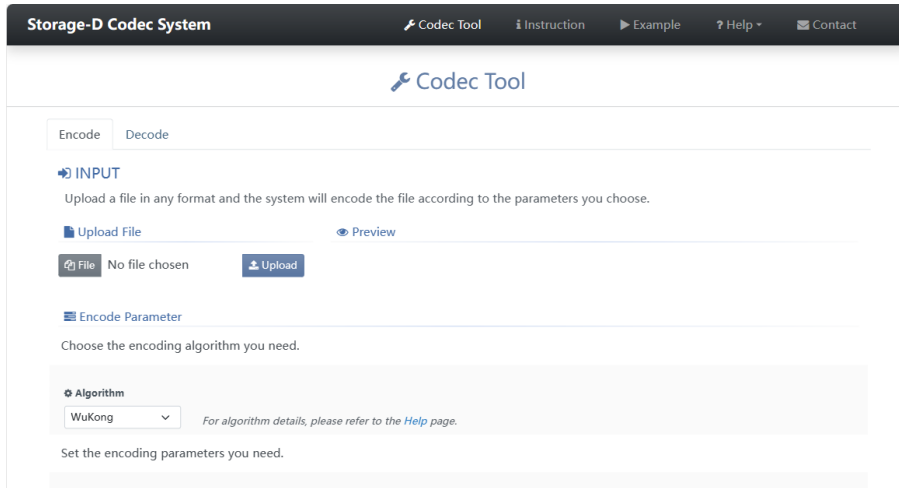
²中国农业科学院深圳农业基因组研究所, 岭南现代农业科学与技术广东省实验室, 农业部农业基因数据分析重点实验室

³华南理工大学生物科学与工程学院

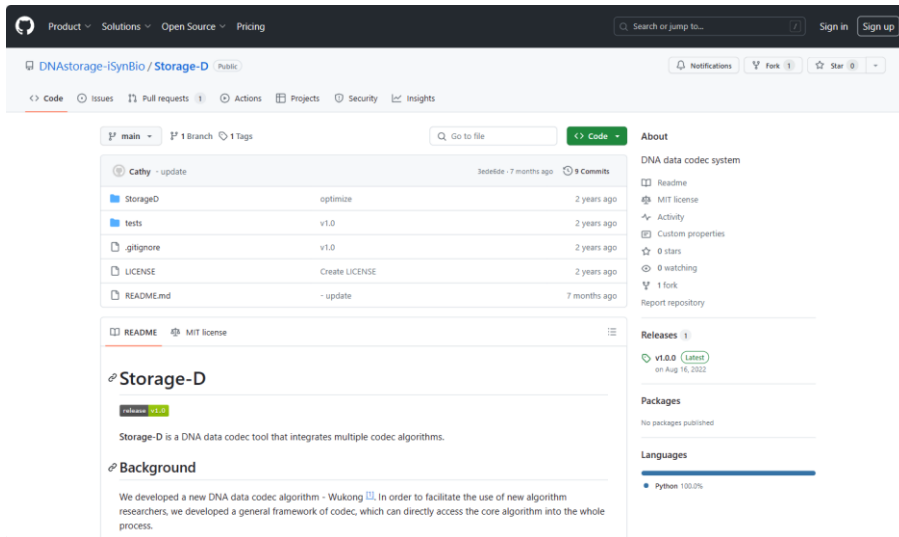


Xiaoluo Huang, Juntong Cui, Wei Qiang, Jianwen Ye, Yu Wang, Xinying Xie, Yuanzhen Li, and Junbiao Dai. 2024. Storage-D: A User-Friendly Platform That Enables Practical and Personalized DNA Data Storage. *iMeta* e168. <https://doi.org/10.1002/imt2.168>

简介



website: <http://storage.dailab.xyz:16666/>

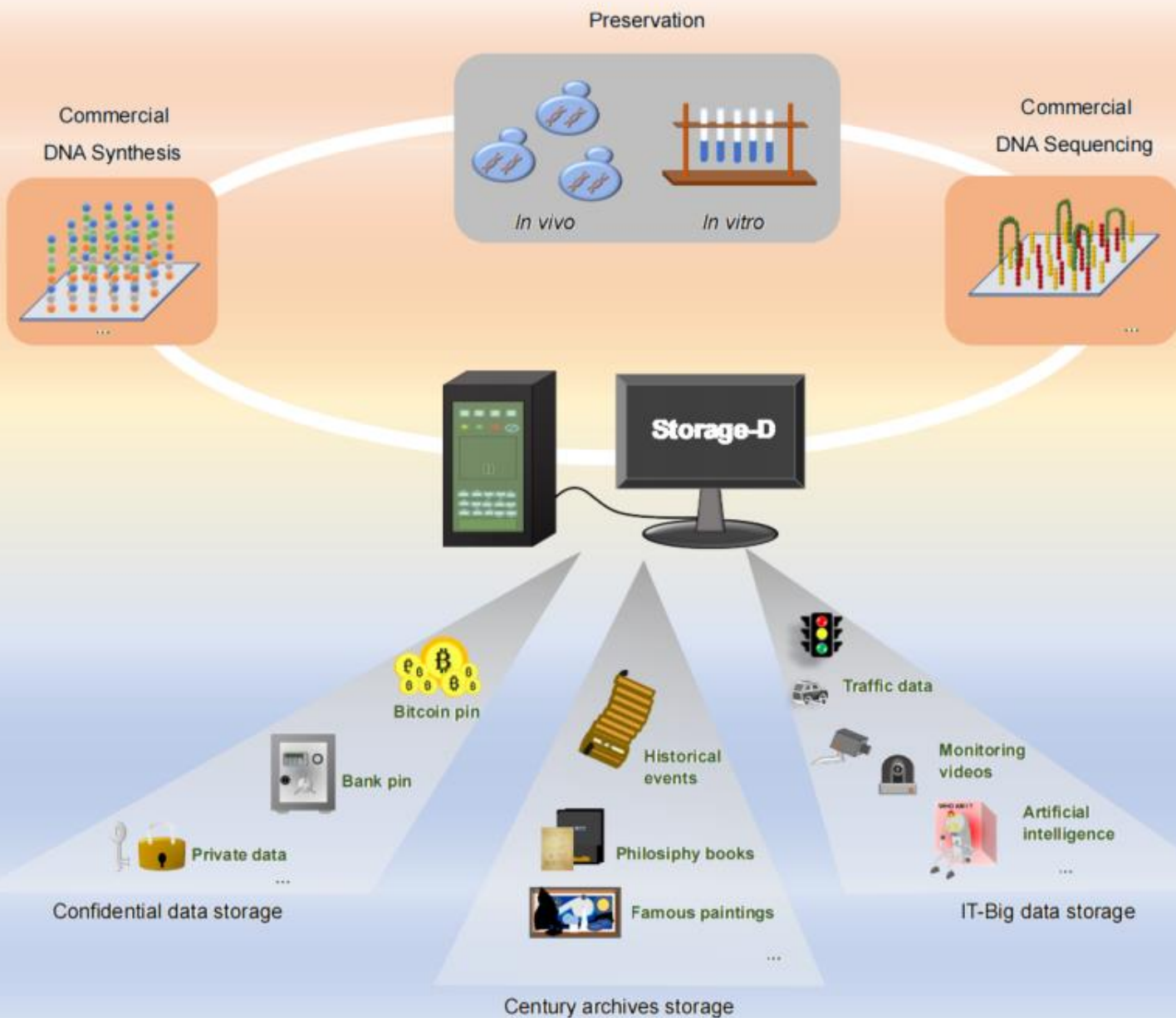


code: <https://github.com/DNASTorage-iSynBio/Storage-D/>

DNA是一种很有前途的数据存储介质，但其实际使用缺乏用户友好的平台。我们开发了基于Web的“Storage-D”编解码平台，将关键的编解码功能模块化，并为实际数据存储提供了个性化选择。平台集成了我们新开发的“悟空”算法，该算法编码逻辑灵活，能够生成大量的编码规则以满足不同的DNA数据存储需求。我们通过体内和体外实验验证了该算法，并提供了网站服务和源代码。

The web server and codes of the platform are available at <http://storage.dailab.xyz:16666/> and <https://github.com/DNASTorage-iSynBio/Storage-D/>, respectively.

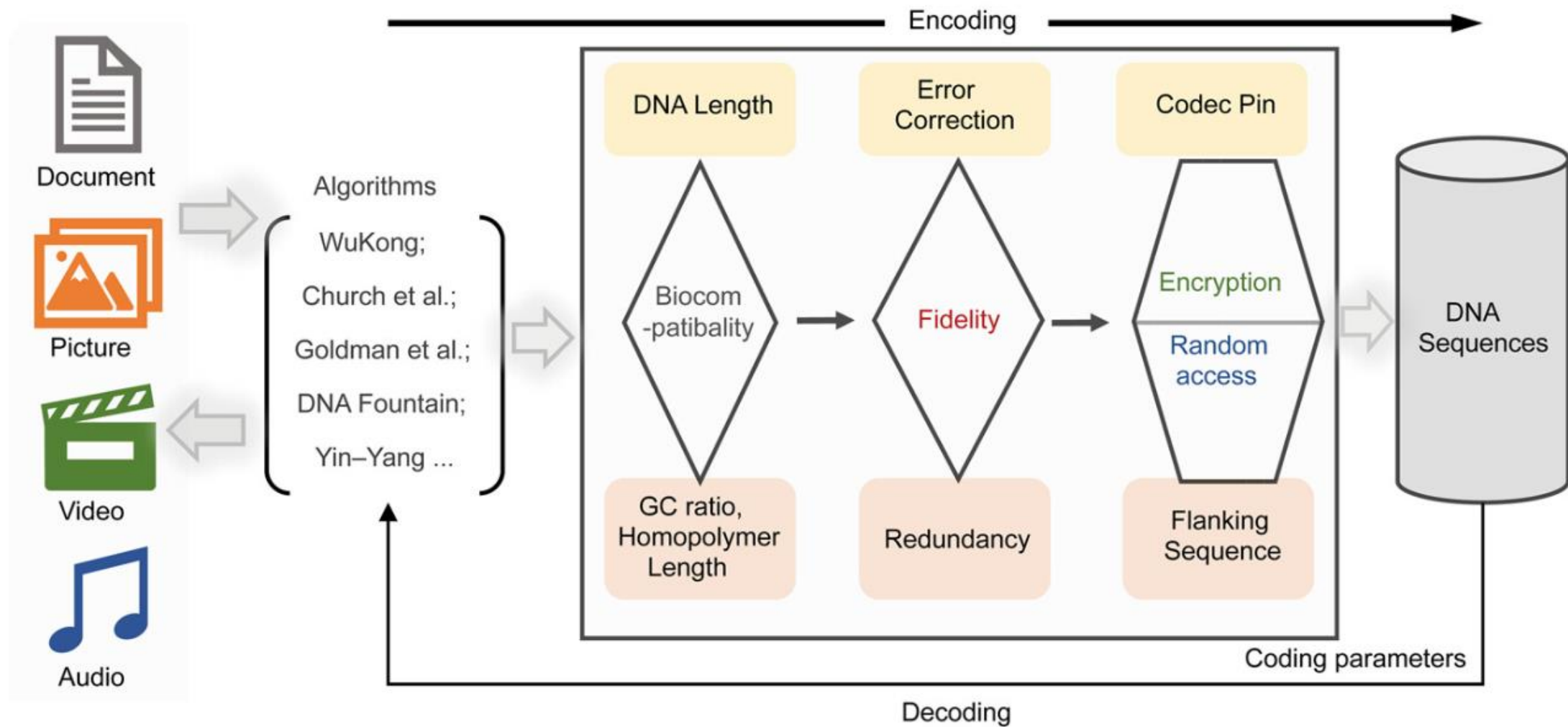
亮点



- DNA数据存储具有密度高、寿命长等显著优势，被认为是应对未来数据存储危机最有前途的介质之一。
- 开发了一个用户友好型平台“Storage-D”，使用户能够通过个性化选择存储任何格式的实用数据。
- 开发了名为“悟空”的新算法，包含大量编解码规则，并具有高度的隐私性。
- “悟空”编码生成的体内及体外DNA存储序列优于早期算法。
- 平台可轻松与其他编解码算法集成，并连接到商业DNA合成和测序平台，实现实用DNA数据存储的完整流程。并提供了网页服务和代码。

结果

Storage-D 的工作流程



结果

Storage-D 网站界面展示

Encode Parameter

Choose the encoding algorithm you need.

Algorithm
WuKong For algorithm details, please refer to the [Help](#) page.

Set the encoding parameters you need.

Encoded Length	Homopolymer	Min GC%	Max GC%
<input type="text" value="200"/>	<input type="text" value="6"/>	<input type="text" value="40"/>	<input type="text" value="60"/>

Codec Pin

**If the [Codec Pin](#) is set to [User Defined](#), remember your [Pin](#) value, because the same value needs to be entered when decoding.*
For parameter details, please refer to the [Help](#) page.

Set the additional parameters you need.

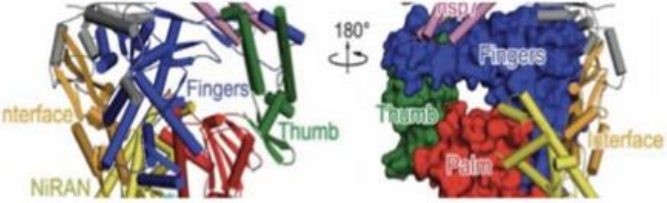
<input checked="" type="checkbox"/> ECC (RS)	Flanking Sequence	<input type="checkbox"/> Redundancy
<input type="text" value="0"/>	<input type="text" value="No"/>	

**ECC (RS): The number of RS Code bytes in each maximum 256 bytes. For details, please refer to [reed-solo](#).*
For parameter details, please refer to the [Help](#) page.

You can fill in an email to receive the encoded result file.

Preview

Ribbon diagram of COVID-19 Virus RNA-dependent RNA polymerase.jpg



OUTPUT
The coding results will be shown here.

Result Preview
Ribbon diagram of COVID-19 Virus RNA-dependent RNA polymerase_wukong.fasta

```
>totalBit:676816, binSegLen:187, leftPrimer:, rightPrimer:, fileExtension:.jpg, bR
>seq_1
CGCGTCGGCCAAGTGCCAGTGAGTGAATGCTCAGCAGGTTGGCAACGGCGGCTTTGAGCATTGATTAATAAG
>seq_2
```

Result Information

Input
File Name: Ribbon diagram of COVID-19 Virus RNA-dependent RNA polymerase.jpg
File size: 676816 (bit)

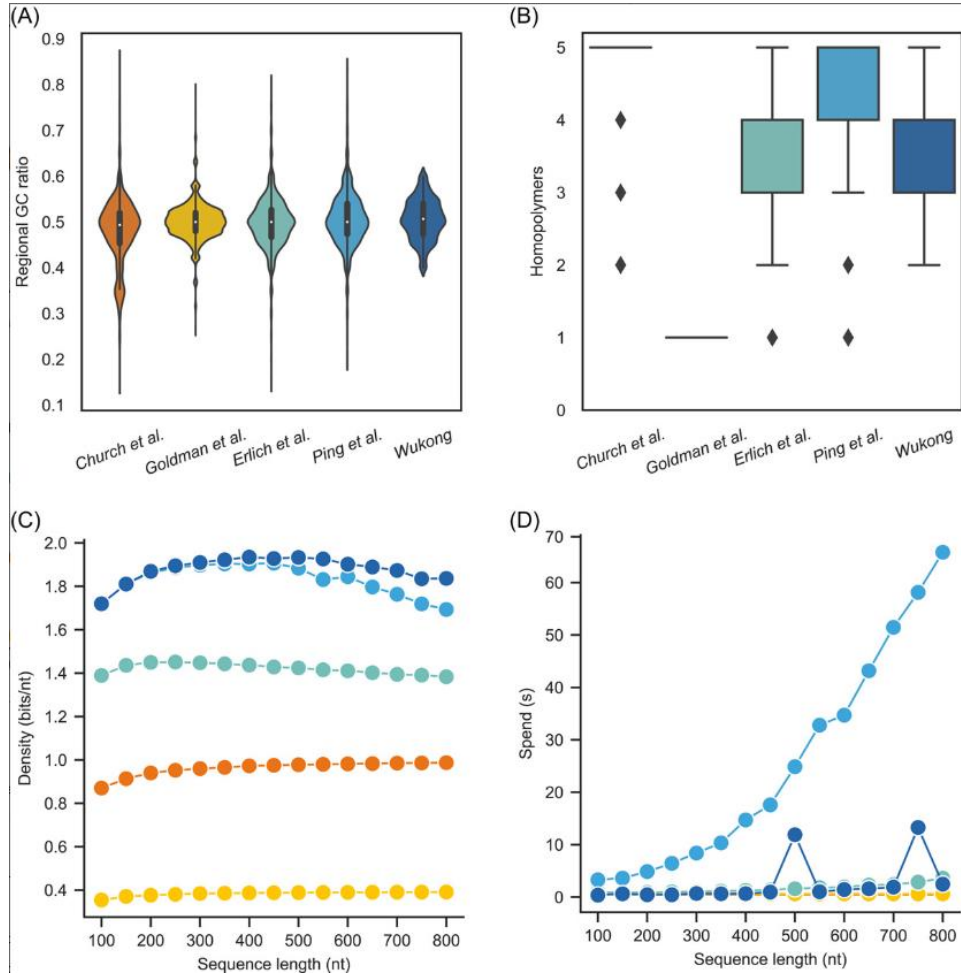
Result
Run Time: 0:00:00.298942
Total Base: 362000 (nt)

[Download](#)

结果

Storage-D 中实现的“悟空”编解码算法

➤ 使用不同算法编码

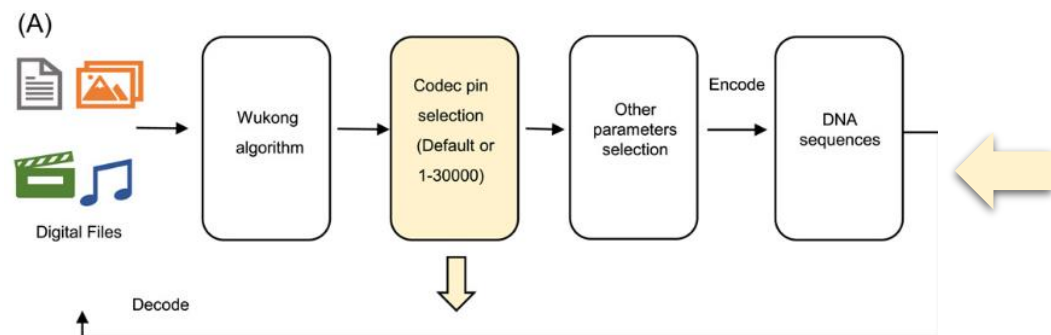


➤ “悟空”和其他算法的比较

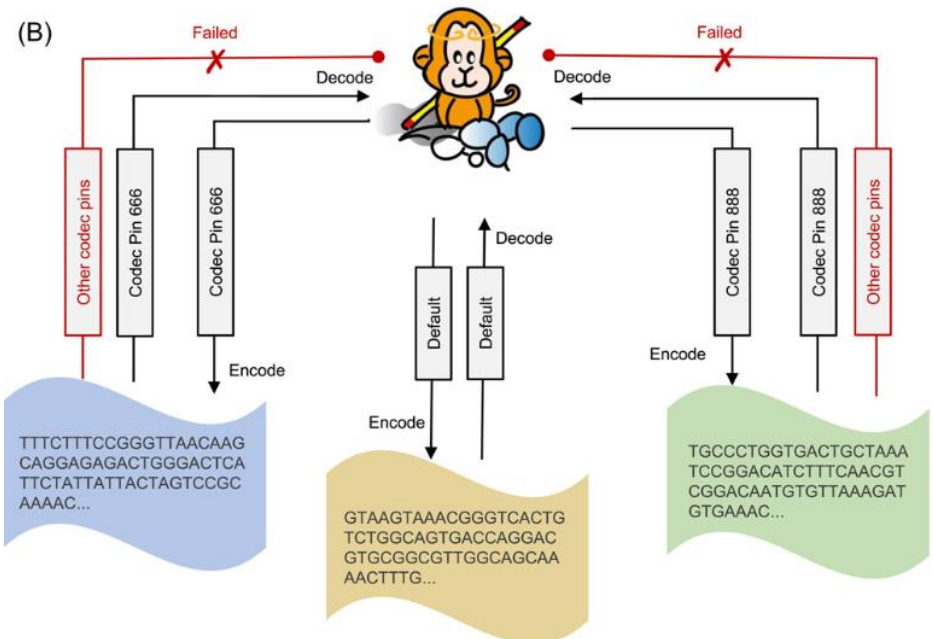
General features	Details	Church <i>et al.</i>	Goldman <i>et al.</i>	Erlich <i>et al.</i>	Ping <i>et al.</i>	Wukong
Density	Coding potential*	1	1.58	1.98	1.95	1.98
Biochemical Compatibility	Regional GC (%)	15–85	26–79	15–80	20–84	40–60 or defined
	Homopolymer	4	1	4	4	4 or defined
Data Fidelity	Error correction	NO	NO	Yes	Yes	Yes or defined
	Redundancy	NO	Yes	Yes	Yes	Yes or defined
Encryption	Codec Pin	--	--	--	1536 [#]	> 20 Trillion
Random-Access	Primer Design	--	--	--	--	Yes or defined

结果

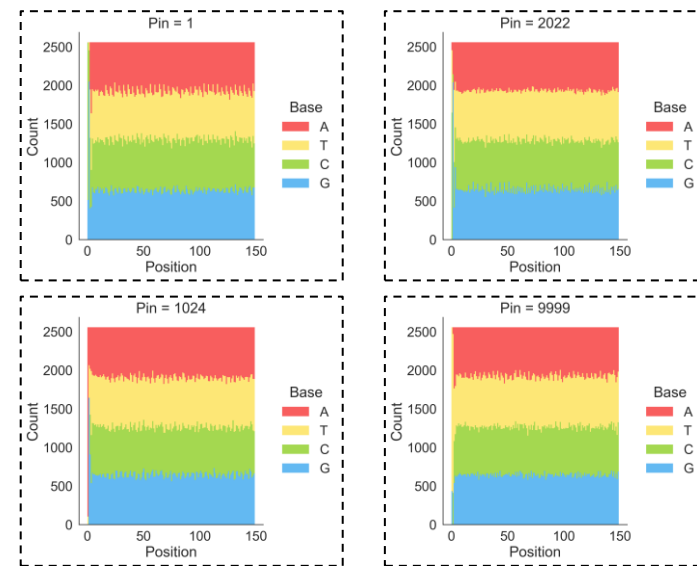
“悟空”算法使用不同密码进行编解码的说明



使用不同的编解码密码对数据进行编码

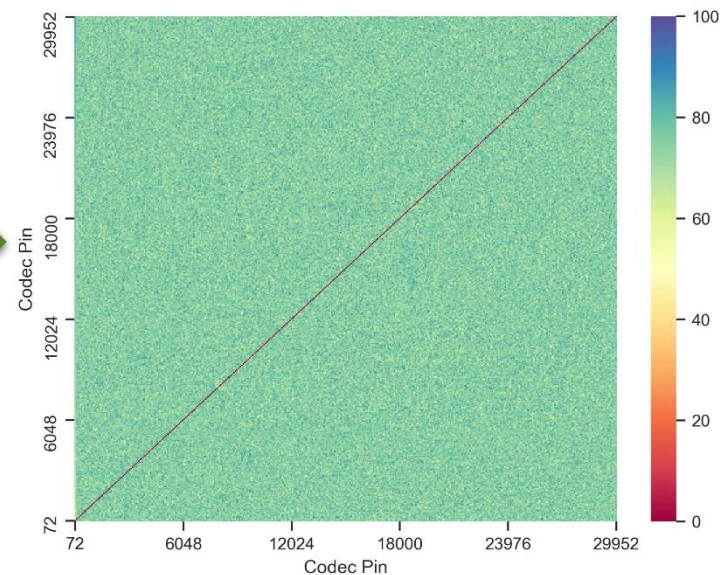


(c)



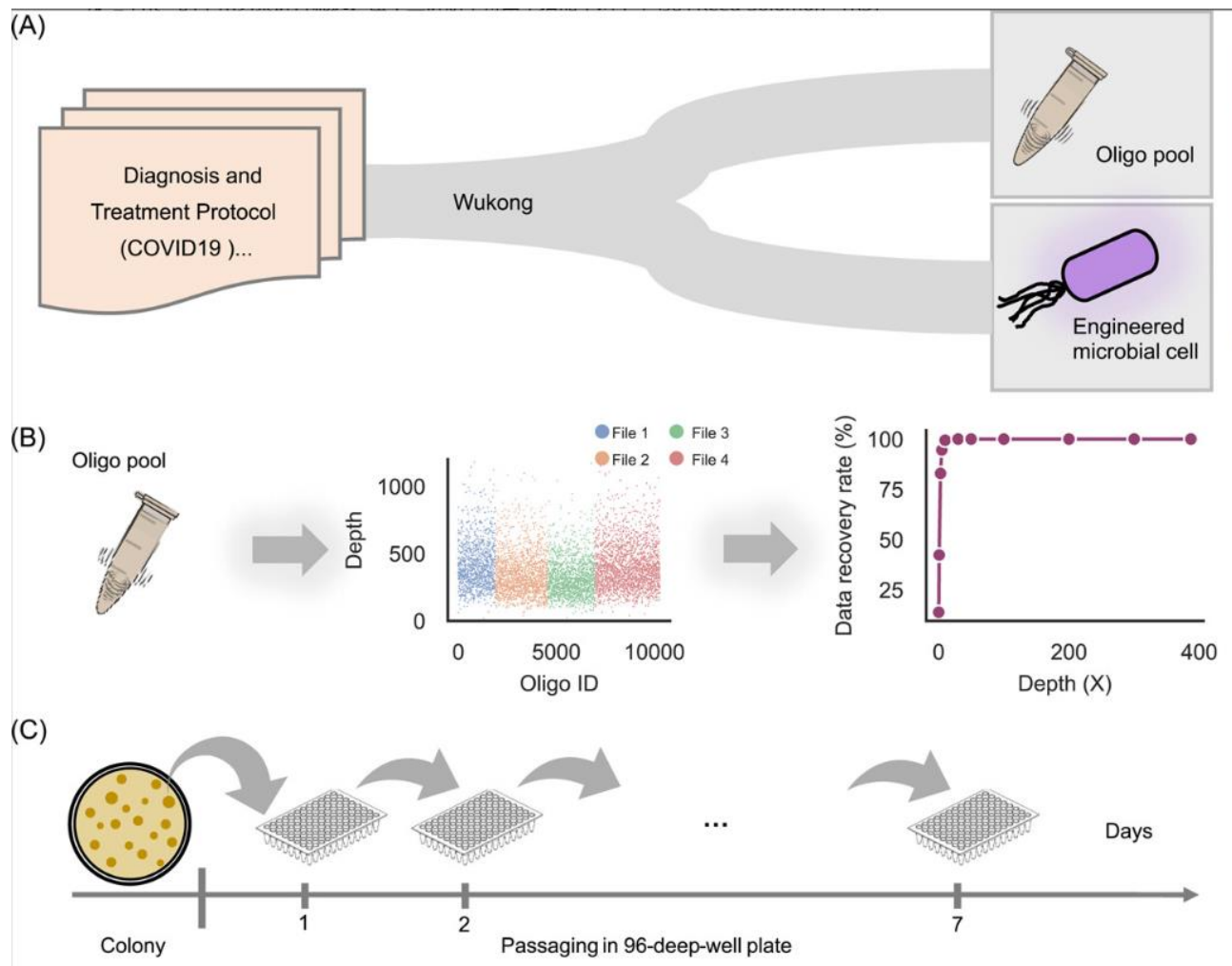
不同编解码密码编码的 DNA 序列之间的汉明距离热图

(D)



结果

通过实际数据存储验证 Storage-D

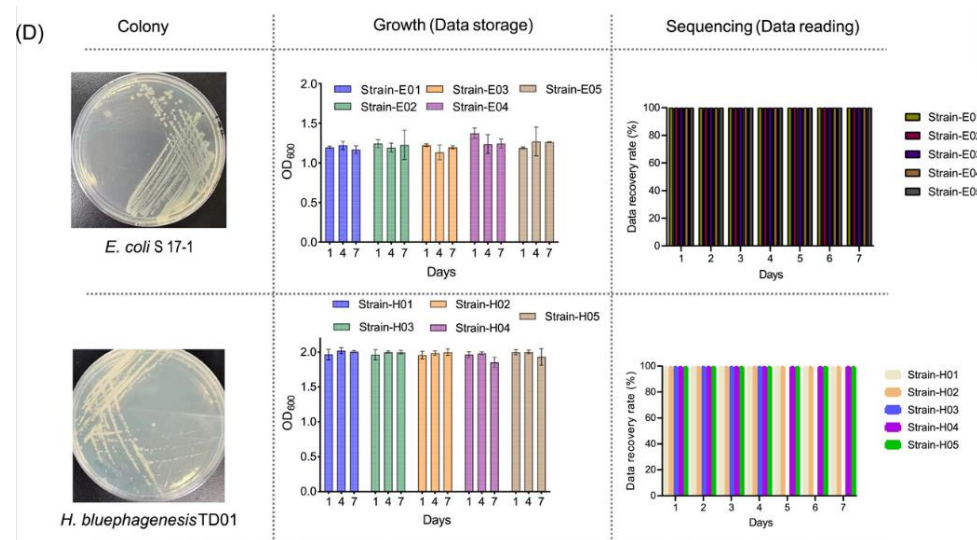


(A) 通过“悟空”算法进行体外和体内数据存储的说明

(B) 通过测序读取寡核苷酸池

(C) 体内数据传递图示

(D) 通过测序获得的生长状态和数据恢复率

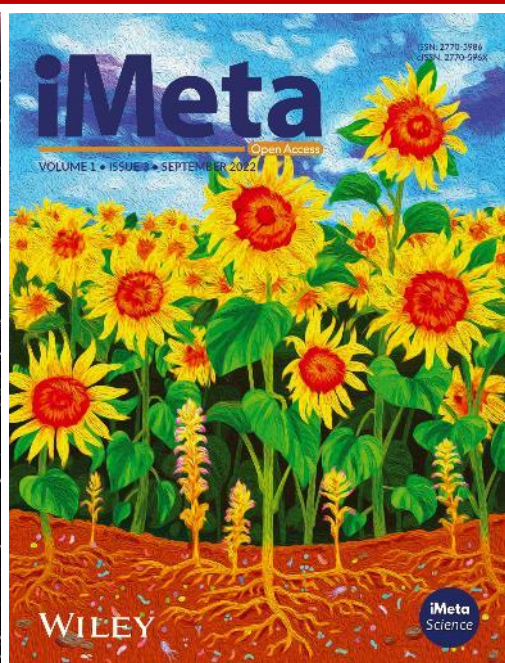




总结

- “Storage-D”将 DNA 数据存储的基本功能模块化，如编解码、纠错和随机访问等，并集成了四种现有算法。
- “悟空”算法可将数据编码为不同的长度、控制均聚物及GC 含量，并可通过设置不同的编解码密码来保障数据的隐私性。通过计算模拟和体内外实验验证了“悟空”算法的可行性。
- 提供了用户友好的在线网页服务以及可供本地下载集成的源代码。

Xiaoluo Huang, Junting Cui, Wei Qiang, Jianwen Ye, Yu Wang, Xinying Xie, Yuanzhen Li, and Junbiao Dai. 2024. Storage-D: A User-Friendly Platform That Enables Practical and Personalized DNA Data Storage. *iMeta* e168. <https://doi.org/10.1002/imt2.168>



“iMeta”是由威立、肠菌分会和本领域数百位华人科学家合作出版的开放获取期刊，主编由中科院微生物所刘双江研究员和荷兰格罗宁根大学傅静远教授共同担任。目的是发表原创研究、方法和综述以促进宏基因组学、微生物组和生物信息学发展。目标是发表前10%(IF > 15)的高影响力论文。期刊特色包括视频投稿、可重复分析、图片打磨、青年编委、前3年免出版费、50万用户的社交媒体宣传等。2022年的三月、六月和九月期已正式在线出版发行!



主页: <http://www.imeta.science>

出版社: <https://wileyonlinelibrary.com/journal/imeta>



投稿: <https://mc.manuscriptcentral.com/imeta>



office@imeta.science



[iMeta](#)

[宣传片](#)

