GENETIC ENGINEERING THROUGH QUANTUM CIRCUITS: CONSTRUCTION OF CODES AND ANALYSIS OF GENETIC ELEMENTS BIOBLOQU

Patrícia Verdugo Pascoal[1], Deborah Bambil[1], Luisa Mayumi Arake de Tacca[1], Rayane Nunes Lima[1], Marco Antônio de Oliveira[1], Mendeli Henning Vainstein[2], Elibio Rech[1]

[1]Embrapa Genetic Resources and Biotechnology/National Institute of Science and Technology in Engineering Biological Systems, ZIP 70770-917, Brasília, Distrito Federal, Brazil.

[2]Institute of Physics, Federal University of Rio Grande do Sul, ZIP 91501-970, Porto Alegre, Rio Grande do Sul, Brazil.

Abstract

The accelerated exploration and engineering of nucleotide sequences are directed towards quantum mechanics and their intrinsic entanglements, implementing the qubits states, including the development of algorithms. The production rate of biological sequencing data has increased to approximately 1 Gb/h, but the ability to analyse these data has not kept pace due to complexity issues and the limitations of classic computing. Despite its own challenges, quantum computing offers a potential solution for analysing biological data and extracting relevant information. Here, we developed and implemented a quantum algorithm that drives searches for genetic information. Our results demonstrated that two codes written based on quantum computing language could precisely search for a target sequence with 50 nucleotides in a DNA sequence database with up to 3022 nucleotides, building a synthetic structure called Quantum Biological Blocks (BioBloQu) inside a minimal JCVI-syn3.0 cells. Our algorithm constitutes a unique starting point for developing a more sophisticated model for the manipulation of data. This tool could be exploited for quantum-enhanced design, building synthetic genomes with desirable traits from the bottom up and exploiting genetic big data. Enhancing the algorithm and expanding device availability are significant targets for accelerating searches and gathering specific biodiversity features for research and applications.

Introduction

The intersection of quantum theory and biology has intrigued researchers for over a century, suggesting a fundamental limit to our understanding of life through the lens of atomic structures.[1] The advent of quantum mechanical computers in the early 1980s[2] set the stage for ground-breaking advances, culminating in the late 1980s and early 1990s with pioneering descriptions[3-7] that were aimed at solving problems that had previously been considered intractable with classic algorithms. In 1996, Grover introduced a quantum algorithm that revolutionized database searches, resulting in a remarkable reduction in complexity from O(N) to O($\sqrt{N}$); however, the practical applications of this algorithm remain limited by the operational demands of large databases.

In the coming years, quantum simulations are expected to make significant advances in areas such as quantum chemistry and high-temperature superconductivity. The community is making progress, but the creation of noise-resistant quantum computers that can handle a variety of problems will likely take another 10 to 15 years. The current challenges include protecting the delicate states of quantum bits (qubits) from errors and unwanted environmental disturbances. This gap between the theoretical potential and the current technological capabilities underscores the transformative promise of quantum computing, particularly in biological research[8]. According to Pan[8], realistic expectations regarding advances in quantum computing are needed since large companies are claiming to offer quantum computing services when they are limited to small-scale algorithms.

Quantum computers operate using quantum bits (qubits), which, unlike classic bits, can exist in multiple states simultaneously because of superposition. The phenomenon of entanglement enables interdependencies between qubit states, facilitating faster computations[9-11]. Importantly, quantum computers should not be seen solely as an evolutionary successor of classic systems[12];

instead, they excel at solving specific, complex problems, whereas classic systems remain optimal for other tasks[13].

Quantum computing opens new avenues for biological research that could lead to breakthroughs in medicine[14], biotechnology[15,16], and our understanding of life itself[17] and evolution[18]. Currently, quantum computing approaches, with their intrinsic and unique capabilities, offer promising solutions for the acceleration, prospection and manipulation of nucleotide sequences with the aim of discovering specific and innovative traits in genomics[19-23]. For a broader perspective on quantum computing in biology, several articles have been published[15,24-33]. However, the lack of sufficient quantum algorithms that are tailored to computational biology challenges hinders the potential of quantum computing to accelerate genetic data processing and manipulation processes in this field[9,10,12,13,34-44].

A quantum search conducted using Grover's algorithm showed the explicit process of constructing a gate, which was programmed using Qibo for N = 8 and M = 2. While the classic algorithm required N + M queries in the worst case, the quantum algorithm found the solution with only one or two queries, demonstrating an improvement in execution time. However, a limitation of the algorithm is that N must be significantly greater than M. In addition, studies have included generalized the algorithm to the fourth dimension to search for codons in DNA sequences using quarts (4-dimensional qudits) and modified the alphabet to include A, C, G, and U[37]. In this context, efforts to develop new quantum algorithms form a key challenge in quantum information technology, which, in turn, demands the combined expertise of the quantum physics and computational biology communities[26].

Owing to hardware requirements, our study was performed on a 5-qubit IBM cloud machine and was limited to a subset of a gene sequence containing four nucleotides[12]. We developed and

implemented two codes that apply quantum algorithms to optimize large-scale genetic sequence analyses using 127 qubits derived from the IBM simulator. The present code contributes to the implementation of a novel quantum computing algorithm for a specific DNA sequence (the "target sequence") in a genetic sequence database and builds a functional “bioblock” that we call BioBloQu in a region of the minimal cell genome (JCVI-syn3). In this work, we extended the applicability of the Grover algorithm by developing and implementing an algorithm that is capable of handling more complex problems, such as searching DNA nucleotide sequences (A, T, G, and C), instead of binary symbols. The diffusion operator remained unchanged, but the oracle was modified to adapt to and operate on ququarts, enabling the algorithm to be efficiently implemented and conduct searches within a larger and more complex dataset of synthetic structures.

Results

Searching for the sequence of interest using QuBio.py code

The primary objective of the quantum simulation conducted using *Qubio.py* was to evaluate the effectiveness of Grover's algorithm in terms of identifying specific DNA sequences with 50 nucleotides (target sequences) within a genetic database (3022 nucleotides). The target DNA sequence was selected based on its relevance to ongoing gene identification research; in this case, a conserved functional domain of the Cas9-like nucleotide sequence found in Brazilian biomes was employed.

This quantum code performs nucleotide sequence comparisons, allowing a mismatch rate of up to 30%. This approach differs from BLAST[45], which is a widely utilized tool for sequence alignment that, although efficient, is predominantly classical. It allows up to 30% differences between the compared sequences, which can be useful in contexts where sequence variations are expected, such as in evolution studies or analyses of homologous genes. Although BLASTN also

identifies similar sequences, the way it calculates similarity values and addresses mismatches is more complex, considering factors such as gap penalties and different forms of similarity.

Focused on applications where speed and tolerance to variations are paramount, the utilized method can be particularly useful in projects that require rapid sequence analyses with a focus on general similarities.

Our code represents an interesting advance regarding the application of quantum computing to sequence comparisons, allowing for fast and flexible analyses in contexts where tolerance to mismatches is desirable. This innovative approach can complement the classic tools, offering new perspectives for genomic research.

The simulation and implementation stages were performed on the IBM Quantum Experience platform, which used a quantum processor with 127 qubits. The Grover's algorithm-based code was configured with a search space corresponding to a genetic database sequence consisting of 3022 nucleotides from a Brazilian metagenome sequence database. On the IBM simulator, the algorithm completed in 1 second, whereas on the real IBM quantum computer, it took slightly longer, at 6 seconds with 1 iteration. This difference was expected because of the additional overhead and complexity involved in executing the algorithm on physical quantum hardware relative to a simulated environment. In the IBM simulator, the numbers of tested iterations were 1, 10, 20, 50 and 100, which ensured that the probability of observing the correct result was maximized and provided a robust comparison. Overall, the real quantum computer was slightly slower but still efficiently handled the same sequence size and number of shots as the simulator did, indicating its practical applicability for use in such bioinformatics tasks. The results of the tests are summarized in Table 1.

Additionally, by performing tests with different numbers of iterations (Fig. 1), it was possible to verify an oscillation pattern in the final probability of the sequence being found in the database.

With one, 10, 20, 50 and 100 iterations, the oscillation in the probability continued in the pattern between 0.26 and 0.28 (26-28%) probabilities of the sequence being found by the code based on Grover's algorithm.

We also tested how the code behaved by changing only 15 nucleotides in the target sequence (30% mismatch rate) and not modifying them in the database. The results indicate that the algorithm was sensitive to this specific change. When we changed more than 15 nucleotides, we observed a sequence finding probability equal to zero (Fig. 2).

Additionally, we performed a simulation to compare the predicted processing times using a quantum computer with various qubit counts (150, 200, 250, and 300 qubits) and a traditional server with 100 threads. These predictions were made for target sequences consisting of 100, 200, and 300 base pairs via the real processing times obtained from a 100-qubit quantum computer and a 100-thread server with 50 base pairs as references. The search was conducted on a 1-petabyte database with a reference sequence of 3 kb. The results indicate that, as expected, increasing the number of qubits significantly reduced the processing time. For example, processing a sequence of 100 base pairs required 0.68 hours with 100 qubits, whereas the same task was completed in only 0.17 hours with 300 qubits. Similarly, for a 300-bp sequence, the time requirement decreased from 2.04 hours with 100 qubits to 0.51 hours with 300 qubits. In comparison, the server with 100 threads required 0.448, 1.024, and 1.6 hours to process 100, 200, and 300 base pairs, respectively (Fig. 3).

These results suggest that quantum computing offers substantial processing efficiency improvements, particularly as the sequence length increases, making it a promising solution for handling large-scale metagenomic data.

Finding a scar in the minimal M. mycoides JCVI-Syn3 cell and building functional biological blocks

New genes identified in metagenome studies, as exemplified above with Cas9 from Brazilian biomes, can be sources of new functions and attributes for genetically modified or even synthetic organisms. Therefore, quantum computing-based tools that assemble genes and detect optimal insertion sites will be valuable techniques for guiding rational genome editing processes or even the *de novo* design of completely synthetic genomes. In that context, we developed a second code based on Grover's algorithm to execute these tasks.

We worked with sequences acquired from the minimized *M. mycoides* JCVI-Syn3 genome given its importance as one of the main landmarks in synthetic biology. The first breakthrough regarding the development of this minimal cell was the creation of *M. mycoides* JCVI-Syn1.0, which was the first cell to be controlled by a genome that was completely chemically synthesized in the laboratory[46]. This strain then served as a chassis for studies in the identification of essential genes and the removal of nonessential genes to generate a minimized version of the original genome, resulting in *M. mycoides* JCVI-Syn3, which included less than 50% of its parental genome[17].

Our second code was first used to identify the unique sequence present in Syn-3B and resulting from the deletion of the nonessential gene *mmsyn1_0531*, which was originally present between the essential genes *mmsyn1_0530* and *rrl* in Syn-1.0. This new gene organization scheme generated a new sequence that can be seen as a scar marking the deletion site of a nonessential gene (Fig. 4). This region could be an interesting insertion site for an exogenous gene, since there

was no risk of disrupting an existing transcriptional unit. Additionally, given their originally distant positions, suggesting independent regulation for each gene, a reduced degree of structural interference with gene expression was expected.

Following the identification of a good insertion site, the code then assembled a genetic structure with important components to create a transcriptional unit (with a promoter, an RBS, a protein sequence and a terminator). The assembled structure was named a BioBloQu, with each individual part described as geneQuantum (Fig. 5).

The generation of biological blocks for the construction process, which we denote as the BioBloQu of a synthetic sequence of interest using the search performed in step 1 with QuBio.py, included the addition of a promoter sequence, an RBS, the sequence of the enzyme of interest and a terminator (which was able to add repressor and enhancer sequences).

This is a visual representation of the power of quantum computing for constructing genetic blocks inspired by BioBricks^{TM}. This approach not only streamlines the design process but also enhances the potential for creating innovative biological systems. As quantum technologies continue to advance, their applications in synthetic biology promise to revolutionize our ability to engineer life and tackle complex biological challenges.

This new genetic structure replaced the scar that was initially identified by the code, generating a modified minimal cell with the capacity to possibly secrete the protein that was present in the designed BioBloQu (Fig. 6).

The addition of these components suggests an effort to modify or improve the functionality of the genome, possibly aimed at engineering new phenotypic traits or optimizing metabolic processes. In this way, quantum computing and the construction and optimization of genetic sequences are possible. This approach enables the exploration of genetic combinations that are

187 computationally infeasible with the traditional techniques, highlighting the intersection of
188 biotechnology with quantum computing. This fact illustrates a transformation in the genome of
189 *Mycoplasma mycoides* and emphasizes the potential of methods based on quantum computing for
190 use in genetic engineering scenarios. It serves as an example of how synthetic biology can be
191 enhanced by emerging technologies, offering new possibilities for research and applications in
192 diverse fields, such as agriculture and biofuels.

193 **Discussion**

194 Some biomedical scientists are exploring how quantum computing can enhance algorithms and
195 machine learning methods in various biological fields, such as protein design. As biologists adopt
196 this technology, they seek to understand its underlying mechanisms and determine when classic
197 computers are sufficient[22].

198 Several approaches have been proposed to optimize DNA sequence searches, from classic
199 algorithms based on string comparisons to preliminary quantum implementations that seek to
200 exploit quantum superposition and interference. However, gaps remain regarding the practical
201 applications of these methods in large genomic databases.

202 Estimating the universal prior distribution is essential for inferring the algorithmic structure of
203 data and discovering causal generative models. Although these metrics are not computable, they
204 can be approximated by limiting the time and memory resources consumed by the utilized
205 computational model. However, the exponential scaling of potential automata makes them
206 intractable on classic computers, particularly for complex cases. Owing to their unstructured
207 outputs and computational irreducibility, traditional heuristics cannot approximate these
208 computations. A quantum circuit framework was proposed in this study to estimate the universal
209 distribution by simulating the superposition of programs for resource-bounded automata. This

approach offers a quantum resource complexity level with linear complexity concerning the data size, achieving a polynomial speedup over the classic exhaustive methods. Thus, the exploration of program-output relationships represents a promising avenue for accelerating quantum searches, with specific data properties inferred from the universal distribution in the quantum superposition task[28].

De[39] introduced new formulations for RNA structures by utilizing quantum pattern recognition and the quantum Hamming distance within Grover's framework, extracting linear regions from these complex forms. To maximize the achievable quantum speedup, methods have been developed for concatenating different linearized segments of RNA structures into longer sequences. The mechanism was simulated using IBM's Qiskit quantum simulator by employing RNA examples acquired from GenBank and mirBase. Iteration complexity comparisons demonstrated the potential for quadratic speedups in quantum computing over the classic methods, and simulation results indicated high search accuracy. These results suggest that future quantum search methods could significantly accelerate the genome sequencing process as quantum computers become more widely adopted. Further research should investigate the effects of decoherence on detection accuracy and assess the computational demands of each iteration on actual quantum hardware.

Kösoglu-Kind et al.[12] examined how genetic information is encoded in nucleotide sequences and identified it through comparative analyses. Variations can occur at the nucleotide level or collectively due to recombination or deletion. Detecting these differences is essential for biology and medicine but requires significant classical computing power due to genomic complexity. To address this issue, the authors proposed the use of flexible representations of quantum images

(FRQIs) for sequence analysis purposes, enabling detailed comparisons and improving the detection accuracy achieved for subtle genetic variations.

The potential of quantum computing to speed up the discovery of specific and innovative traits from big nucleotide sequence data is immense. By providing unparalleled processing power, enhancing the precision of trait discovery, optimizing genetic algorithms, integrating with AI, and enabling real-time analyses, quantum computing could revolutionize how we approach genomics and trait development[47]. As this technology matures, it promises to unlock new possibilities in biotechnology, agriculture, and medicine, paving the way for ground-breaking innovations that could reshape our understanding of life[48,49].

After analysing the comparative simulation of the predicted processing times using a quantum computer and our robust classic computer server with 100 threads (Fig. 3), the results suggested that quantum computing offers substantial improvements in processing efficiency, particularly as the sequence length increases, making it a promising solution for handling large-scale metagenomic data. Therefore, to match our server with 100 processing threads, the quantum computer needed approximately 150 available qubits. IBM's quantum computer currently has 400 qubits, but less than 200 of them are available for use in either simulations or as real quantum units[50]. Even at a smaller scale, quantum computing can provide efficiency advantages as the number of base pairs contained in the target sequence increases. The use of a greater number of qubits (such as 250 or 300) resulted in processing times that were significantly lower than those of classic computing, suggesting that quantum computing may be more suitable for problems requiring high computational power. The representation of classic computing exhibits a linear increase in processing time, indicating that as the volume of data increases, classic computing may become impractical for tasks that demand intensive processing steps. These observations highlight

the potential of quantum computing to transform fields that require the analysis of large amounts of data, such as bioinformatics, molecular modelling, and other scientific applications. Challenges remain, and there is a long way to go before quantum computing becomes established, and we are working in parallel with this progress to maximize its capabilities and expand its achievements within biology and biotechnology.

Here, the use of one iteration is sufficient for the objective and demands less time and computational power, as it takes approximately one second for one iteration and 49 seconds for 100 iterations. Increasing the number of iterations in a quantum search algorithm can lead to a decrease in the probability of finding the target, which can be attributed to several factors that are related to the nature of quantum algorithms.

According to Batra et al.[14], algorithms can potentially process large datasets more efficiently, reducing the time required for drug discovery processes. The use of quantum computing can lead to more accurate molecular property predictions, facilitating the identification of promising drug candidates. However, their paper also had several limitations. The current state of quantum computing hardware is still in its infancy, which limits the practical implementations of quantum machine learning algorithms in real-world drug discovery scenarios.

The study published by Rodríguez[37], which also used the approach employed in our code, exemplified the trade-offs between computational complexity and algorithmic effectiveness, since the practical results obtained here (100 iterations, 0.28 probability, and 48 seconds) demonstrate such complexity. This analysis can be useful for future optimizations, enabling the exploration of different configurations of quantum circuits to improve the efficiency of the algorithm in similar contexts[31,51].

Studies attempting to generate biological blocks using quantum computing have not yet been reported, and the present study aimed to explore this possibility because quantum computing and synthetic biology are on the rise. Furthermore, the need to quickly and effectively explore large datasets to accelerate and sensitize the construction of biological building blocks is of great interest in the context of biotechnology, especially genetic and biological systems engineering.

The discussion presented here encapsulates a promising and innovative frontier at the intersection of quantum computing and biological sciences, particularly in the domains of genomics and genetic engineering. As researchers increasingly explore the applications of quantum algorithms for solving complex biological problems, it becomes clear that significant advances can be achieved, thereby addressing long-standing challenges in these fields. First, the potential of quantum computing to improve DNA sequence analysis and protein design algorithms cannot be overstated. The ability to leverage quantum superposition and interference offers a paradigm shift in terms of how we process biological data. The traditional computational methods often struggle with the exponential complexity of genomic information, making them unsuitable for large-scale analyses. Owing to its inherent ability to simultaneously process vast datasets, quantum computing presents a viable solution to these challenges. The existing frameworks, such as the quantum circuit approach for estimating universal distributions and the use of quantum pattern recognition in RNA structure analyses, exemplify the innovative strategies that are being developed to effectively harness this technology.

Furthermore, the findings derived from recent studies, including those published by De[39] and Kösoglu-Kind et al.[12], highlight the feasibility of achieving substantial speedups in genetic analyses through quantum computing. The potential for attaining quadratic and polynomial speedups relative to the classic methods suggests that quantum technologies could significantly

reduce the time required for genome sequencing and variant detection. This speedup is particularly crucial as the volume of genomic data continues to grow exponentially, driven by advances in sequencing technologies and the growing interest in personalized medicine.

Here, our study contributes to the small-scale assessment of the possibility of accelerating applied analyses with real genomic data and the perspective of advancing quantum computing in the biological sciences. It is a way to not only keep up with the current hype but also to follow the evolution of a science that, despite being an emerging field, is extremely important for the large volume of biological data that will need to be processed in the future.

In addition, the integration of quantum computing with artificial intelligence represents an attractive avenue for future research. By optimizing genetic algorithms and facilitating real-time data analyses, quantum computing can revolutionize trait discovery processes in agriculture and biotechnology. The potential to accelerate the identification of beneficial traits and increase crop resilience can generate significant benefits in terms of food security and sustainability. As the world faces urgent challenges such as climate change and population growth, the ability to innovate agricultural practices through quantum-enabled solutions has become increasingly vital.

Despite these promising advances, it is essential to recognize the challenges that remain regarding the practical implementation of quantum computing in real-world biological applications. The current state of quantum hardware is still evolving, and qubit availability and coherence time limitations represent significant obstacles. As researchers work to develop more robust and scalable quantum systems, the path to practical applications in biology and genetic engineering will become clearer.

In conclusion, this research represents a significant contribution to the field of applying quantum computing to biology and genetic engineering. The innovative approaches discussed herein

provide a foundation for further exploration and experimentation, illuminating the potential of quantum technologies to transform our understanding of complex biological systems. As we advance in this exciting domain, it is imperative that we continue to foster interdisciplinary collaboration, bridging the gap between quantum computing and the biological sciences. Such synergy will undoubtedly lead to ground-breaking discoveries and innovations, reshaping the biotechnology landscape and offering new solutions to some of the most pressing challenges in healthcare, agriculture and beyond. The future of quantum computing in biology is immensely promising, and continued explorations of its capabilities will be the key to unlocking new frontiers in science and technology.

References

1. Bohr, N. Biology and quantum theory. *Nature* **129**, 343–343 (1931).
2. Benioff, P. The computer as a physical system: a microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. *J. Stat. Phys.* **22**, 563–591 (1980).
3. Deutsch, D. Quantum theory, the Church–Turing principle and the universal quantum computer. *Proc. R. Soc. Lond. A. Math. Phys. Sci.* **400**, 97–117 (1985).
4. Bernstein, E. & Vazirani, U. Quantum complexity theory in *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing - STOC '93* 11–20 (ACM Press, New York, NY, USA, 1993).
5. Yao, A. C. C. Quantum circuit complexity in *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science* 352–361 (IEEE, Palo Alto, CA, USA, 1993).
6. Berthiaume, A. & Brassard, G. Oracle quantum computing. *J. Mod. Opt.* **41**, 2521–2535 (1994).

7. Shor, P. W. Algorithms for quantum computation: discrete logarithms and factoring in *Proceedings 35th Annual Symposium on Foundations of Computer Science* 124–134 (IEEE, Santa Fe, NM, USA, 1994).

8. Pan, J. W. Quantum technologies need big investments to deliver on their big promises. *Nature* **638**, 862 (2025).

9. Gill, S. S. et al. Quantum computing: a taxonomy, systematic review and future directions. *Softw. Pract. Exp.* **52**, 66–114 (2021).

10. Graham, T. M. et al. Multi-qubit entanglement and algorithms on a neutral-atom quantum computer. *Nature* **604**, 457–462 (2022).

11. Shrivastava, P., Soni, K. K. & Rasool, A. Evolution of quantum computing based on Grover's search algorithm in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* 1–6 (IEEE, Kanpur, India, 2019).

12. Kösoglu-Kind, B. et al. A biological sequence comparison algorithm using quantum computers. *Sci. Rep.* **13**, 14552 (2023).

13. Nandhini, S., Singh, H. & Akash, U. N. An extensive review on quantum computers. *Adv. Eng. Softw.* **174**, 103337 (2022).

14. Batra, K. et al. Quantum machine learning algorithms for drug discovery applications. *J. Chem. Inf. Model.* **61**, 2641–2647 (2021).

15. Andersson, M. P., Jones, M. N., Mikkelsen, K. V., You, F. & Mansouri, S. S. Quantum computing for chemical and biomolecular product design. *Curr. Opin. Chem. Eng.* **36**, 100754 (2022).

16. Molinari, M. D. C. et al. Navigating the path from lab to market: regulatory challenges and opportunities for genome editing technologies for agriculture in *Plant Genome Editing Technologies. Interdisciplinary Biotechnological Advances* (eds Chen, J. T. & Ahmar, S.) 25–63 (Springer, Singapore, 2024).

17. Hutchison, C. A. et al. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).

18. Arber, W. Elements for a theory of molecular evolution. *Gene* **317**, 3–11 (2003).

19. Outeiral, C. et al. The prospects of quantum computing in computational molecular biology. *WIREs Comput. Mol. Sci.* **11**, e1481 (2021).

20. Ma, H. et al. Multiscale quantum algorithms for quantum chemistry. *Chem. Sci.* **14**, 3190–3205 (2023).

21. Nofer, M., Bauer, K., Hinz, O., van der Aalst, W. & Weinhardt, C. Quantum computing. *Bus. Inf. Syst. Eng.* **65**, 361–367 (2023).

22. Marx, V. Biology begins to tangle with quantum computing. *Nat. Methods* **18**, 715–719 (2021).

23. Marchetti, L. et al. Correction to: quantum computing algorithms: getting closer to critical problems in computational biology. *Brief. Bioinform.* **24**, bbad134 (2022).

24. Lambert, N. et al. Quantum biology. *Nat. Phys.* **9**, 10–18 (2013).

25. Hollenberg, L. C. L. Fast quantum search algorithms in protein sequence comparisons: Quantum bioinformatics. *Phys. Rev. E* **62**, 7532–7535 (2000).

26. Fedorov, A. K. & Gelfand, M. S. Towards practical applications in quantum computational biology. *Nat. Comput. Sci.* **1**, 114–119 (2021).

27. Nimbe, P., Weyori, B. A. & Adekoya, A. F. Models in quantum computing: a systematic review. *Quantum Inf. Process.* **20**, 80 (2021).

28. Sarkar, A., Al-Ars, Z. & Bertels, K. Estimating algorithmic information using quantum computing for genomics applications. *Appl. Sci.* **11**, 2696 (2021).

29. Roman-Vicharra, C. & Cai, J. J. Quantum gene regulatory networks. *NPJ Quantum Inf.* **9**, 67 (2023).

30. Pal, S., Bhattacharya, M., Lee, S. S. & Chakraborty, C. Quantum computing in the next-generation computational biology landscape: from protein folding to molecular dynamics. *Mol. Biotechnol.* **66**, 163–178 (2024).

31. Pal, S., Bhattacharya, M., Dash, S., Lee, S. S. & Chakraborty, C. Future potential of quantum computing and simulations in biological science. *Mol. Biotechnol.* **66**, 2201–2218 (2024).

32. Aroche, R. R., García, Y. M. O., Arellano, M. A. M. & Leal, A. R. DNA as a perfect quantum computer based on the quantum physics principles. *Scientific reports* **14**, 11636 (2024).

33. Yang, X., Huang, K., Yang, D., Zhao, W. & Zhou, X. Biomedical big data technologies, applications, and challenges for precision medicine: a review. *Glob. Chall.* **8**, 2300163 (2024).

34. Khan, T. M. & Robles-Kelly, A. Machine learning: quantum vs classical. *IEEE Access* **8**, 219275–219294 (2020).

35. Robert, A., Barkoutsos, P. K., Woerner, S. & Tavernelli, I. Resource-efficient quantum algorithm for protein folding. *NPJ Quantum Inf.* **7**, 38 (2021).

36. Efthymiou, S. et al. Qibo: a framework for quantum simulation with hardware acceleration. *Quantum Sci. Technol.* **7**, 015018 (2021).

37. Rodríguez, J. B. *Application of Grover's Quantum Algorithm for String Matching* (Treballs Finals de Grau de Física. Facultat de Fisica, Universitat de Barcelona, 2022).

38. Cordier, B. A., Sawaya, N. P. D., Guerreschi, G. G. & McWeeney, S. K. Biology and medicine in the landscape of quantum advantages. *J. R. Soc. Interface* **19**, 20220541 (2022).

39. De, R. Quantum accelerated pattern matching for genome sequencing in complex RNA secondary structures in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)* 842–843 (IEEE, Broomfield, CO, USA, 2022).

40. Nałęcz-Charkiewicz, K. & Nowak, R. M. Algorithm for DNA sequence assembly by quantum annealing. *BMC Bioinform.* **23**, 122 (2022).

41. Khanal, B., Orduz, J., Rivas, P. & Baker, E. Supercomputing leverages quantum machine learning and Grover's algorithm. *J. Supercomput.* **79**, 6918–6940 (2023).

42. Bhuvaneswari, S., Deepakraj, R., Urooj, S., Sharma, N. & Pathak, N. Computational analysis: unveiling the quantum algorithms for protein analysis and predictions. *IEEE Access* **11**, 94023–94033 (2023).

43. How, M. L. & Cheah, S. M. Business renaissance: opportunities and challenges at the dawn of the quantum computing era. *Businesses* **3**, 585–605 (2023).

44. Schmidt, B. & Hildebrandt, A. From GPUs to AI and quantum: three waves of acceleration in bioinformatics. *Drug Discov. Today* **29**, 103990 (2024).

45. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

435 46. Gibson, D. G. et al. Creation of a bacterial cell controlled by a chemically synthesized
436 genome. *Science* **329**, 52–56 (2010).
437 47. Goswami, S. & Sharma, S. Artificial intelligence, quantum computing and cloud
438 computing enabled personalized medicine in next generation sequencing bioinformatics
439 in *2024 IEEE International Conference on Interdisciplinary Approaches in Technology*
440 *and Management for Social Innovation (IATMSI)* 1–5 (IEEE, Gwalior, India, 2024).
441 48. Mauranyapin, N. P., Terrasson, A. & Bowen, W. P. Quantum biotechnology. *Adv.*
442 *Quantum Technol.* **5**, 2100139 (2022).
443 49. Souza, K. X. S. D. et al. Quantum computing: current and potential applications in digital
444 agriculture. *Pesqui. Agropecu. Bras.* **59**, e03753 (2024).
445 50. Abughanem, M. IBM quantum computers: evolution, performance, and future directions.
446 *J. Supercomput.* **81**, 687 (2025).
447 51. Pyrkov, A. et al. Complexity of life sciences in quantum and AI era. *WIREs Comput.*
448 *Mol. Sci.* **14**, e1701 (2024).
449 52. Allu, S. N., Naidu, T. A. & Rao, K. G. Optimized quantum circuit implementation of
450 SATURNIN for Grover's algorithm. *Int. J. Inf. Technol.* **16**, 2955–2964 (2024).
451 53. Zhang, M. L., Li, H. B. & Jin, Y. Application and perspective of CRISPR/Cas9 genome
452 editing technology in human diseases modeling and gene therapy. *Front. Genet.* **15**,
453 1364742 (2024).
454 54. Grover, L. K. A fast quantum mechanical algorithm for database search in *Proceedings of*
455 *the Twenty-Eighth Annual ACM Symposium on Theory of Computing - STOC '96* 212–
456 219 (ACM Press, New York, NY, USA, 1996).
457

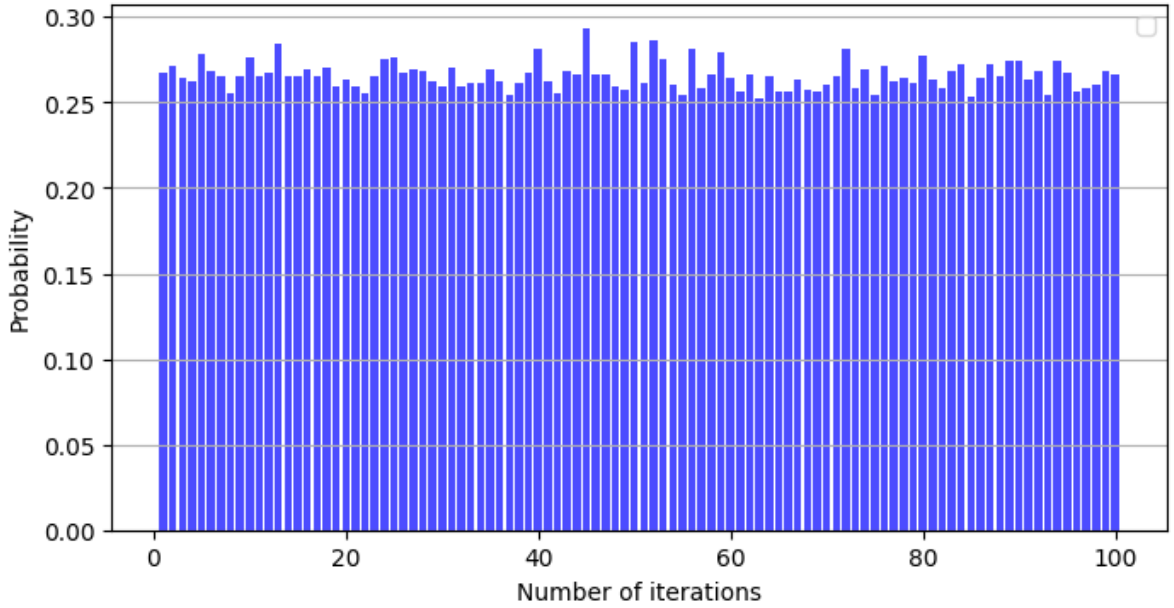
Tables

Table 1. Comparative performance of different computers in terms of executing nucleic acid sequence search algorithms.

	Classical computer/BLASTN*	**IBM Simulator**	**IBM Quantum Computer**
Time of running (seconds)	0.342	4	6
Iterations	*N/A*	1	1
Target_sequence size (nt)	50	50	50
Database_sequence (nt)	3022	3022	3022

*128 CPU Processor.

462 **Figure legends**

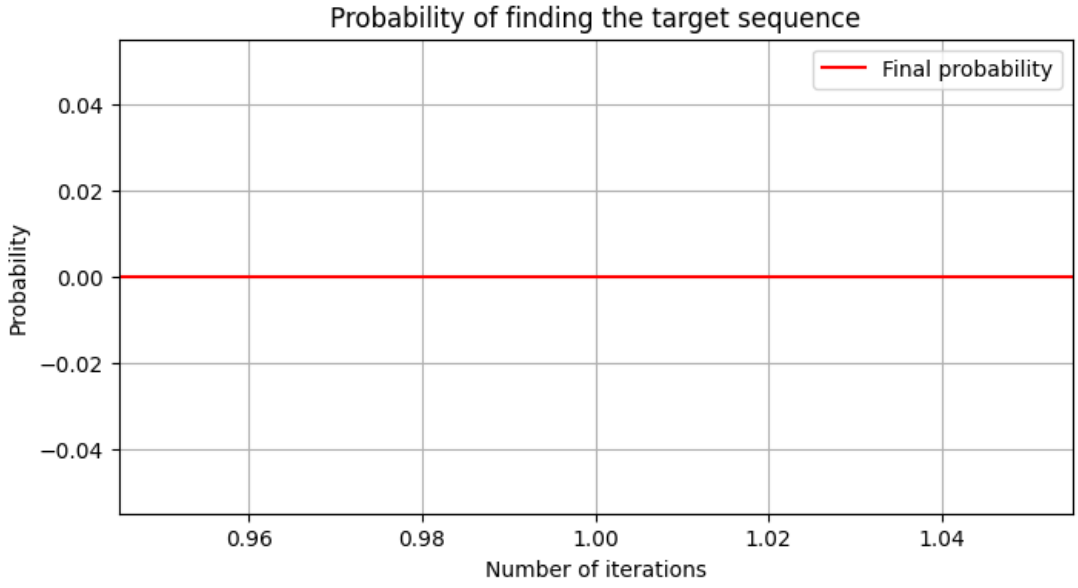


463

464 **Fig. 1**. Probability of finding the target sequence in the database of Brazilian metagenome

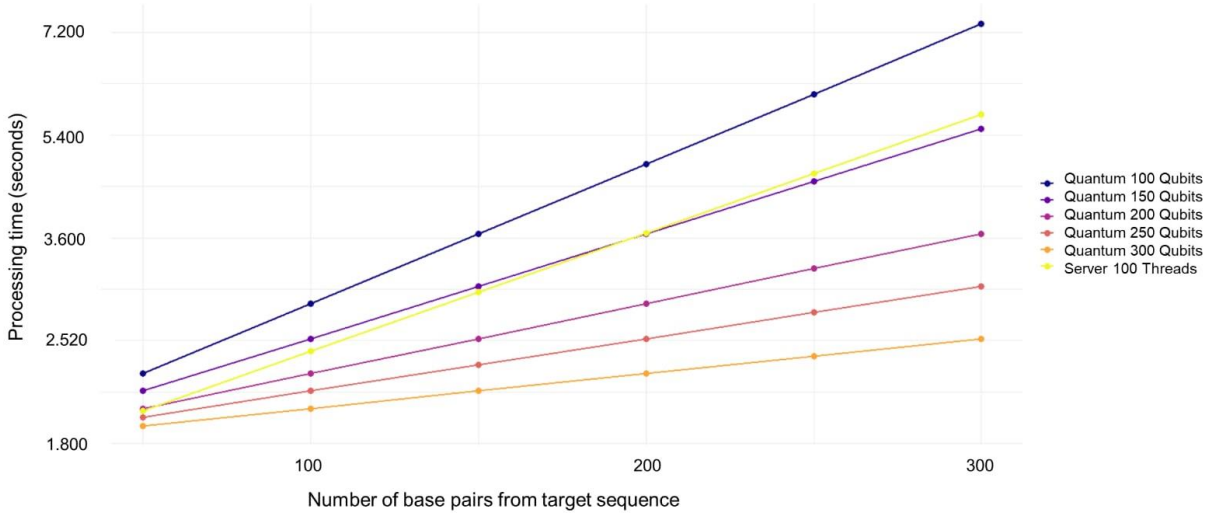
465 sequences according to the QuBio.py code. Running time: 100 iterations ran in 49 seconds using

466 127 qubits of the IBM simulator.



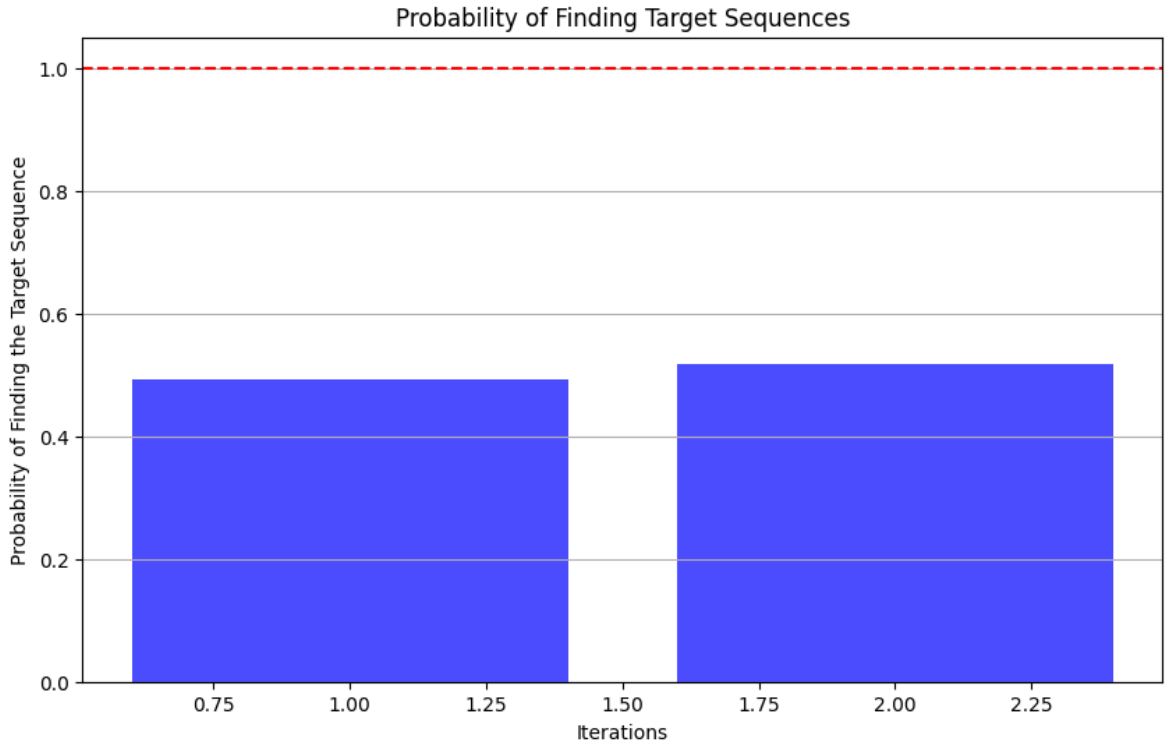
467

468 **Fig. 2.** The final probability of finding a sequence with more than a 30% mismatch rate in the
469 database. The process took 0 seconds on the IBM simulator using 1 iteration.



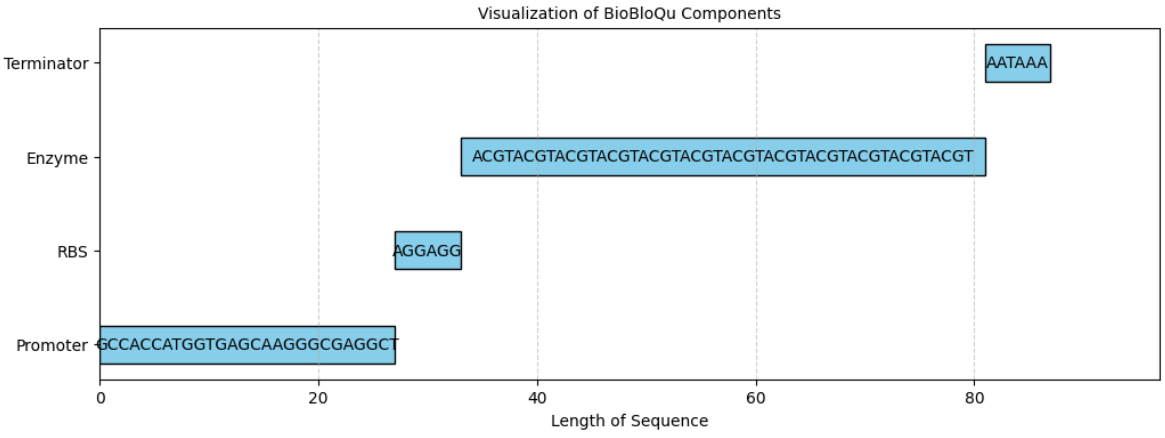
470

471 **Fig. 3.** Comparison among the processing times required for sequence analysis using quantum
472 computers and traditional servers. Processing times needed for target sequences of 100, 200, and
473 300 base pairs using a quantum computer with various qubit counts (100, 150, 200, 250, and 300
474 qubits) and a traditional server with 100 threads.



475

476 **Fig. 4**. Results of the search for two sequences for identifying the scar produced by the removal
477 of an essential gene from the minimized genome of Mycoplasma mycoides (JCVI-Syn3B) using
478 code developed in-house via quantum computing.



479

480 **Fig. 5**. Representation of the structures of the components contained in a BioBloQu.

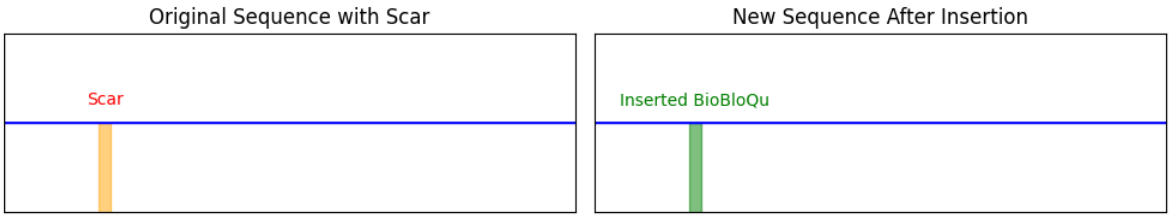


Fig. 6. Representation of the region with the presence of the scar in the genome of Mycoplasma mycoides (JCVI-Syn3B, left) and the subsequent addition of genetic building blocks (called BioBloQu) created via quantum computing (with a promoter, an RBS, a protein/enzyme-encoding gene sequence and a terminator).

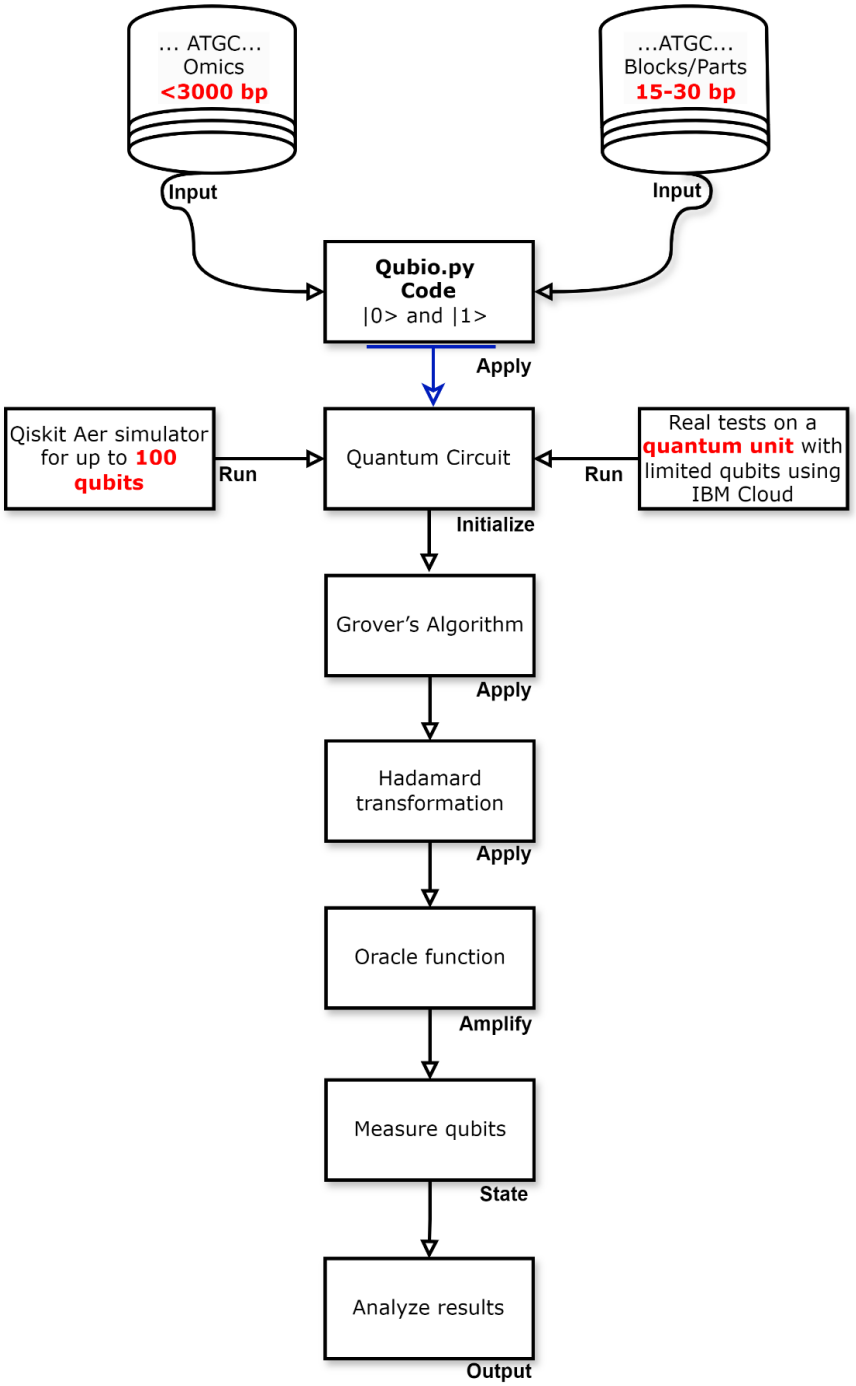
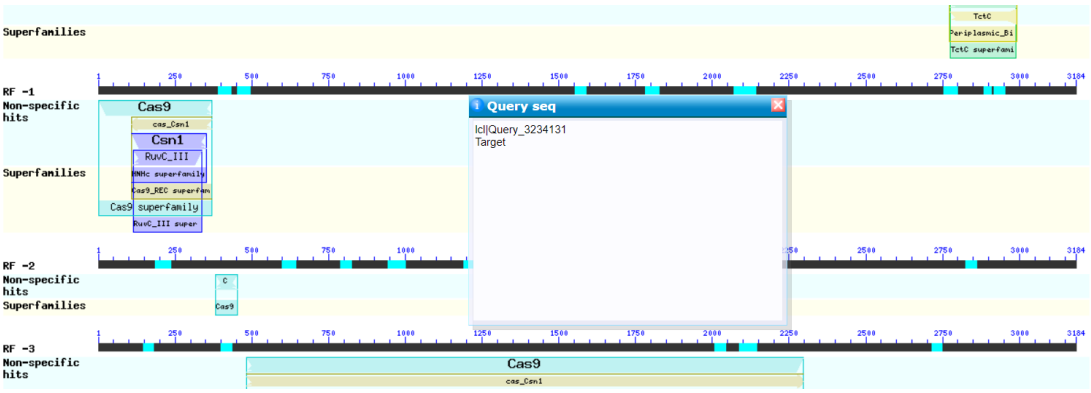


Fig. 7. Flowchart of the Qubio.py program. The Qubio.py program accepts input sequences of both omics data (<3000-bp DNA target) and protein blocks/parts (12 to 50 bp) and initializes a quantum circuit. The processing phase occurs using a Qiskit-based simulator (100 qubits) and real quantum units (IBM Cloud). The core algorithm leverages Grover's algorithm, beginning

491 with the use of the Hadamard transformation to create a superposition of the states, followed by
492 amplitude amplification via the Oracle function. After the qubit states are measured, the results
493 are analysed to identify the target sequence.
494



495

496 **Fig. 8**. The definition of a target sequence based on highlighting the short sequence of the
497 conserved functional domain of a protein.

Methods

The diagram below (Fig. 7) provides a high-level overview of the quantum Qubio.py code developed *in-house* and designed for genetic sequence analyses using Grover's algorithm on a quantum computing cloud platform. This algorithm is well suited for bioinformatics applications, specifically in the context of sequence searches within large databases, because it provides a quadratic speedup over the classic search algorithms, allowing for the handling of genomic data with an immense scale. While the classic search methods require O(N) comparisons, where N is the size of the database, Grover's algorithm reduces this number to $O(\sqrt{N})$[52]. The process begins with two main inputs, the target DNA sequence and the database sequence, where the superposition of quantum states allows multiple comparisons to occur simultaneously. The target DNA sequence is the specific genetic sequence of interest that needs to be analysed. This could be a particular gene or segment related to a specific trait. On the other hand, the database sequence represents a collection of genetic sequences stored in a database. These sequences may come from various sources, such as whole genomes, gene libraries, or databases of genetic variations[12].

Once the data are prepared, the core of the process involves applying Grover's algorithm, which is a quantum algorithm that was designed for efficiently searching unsorted databases. By leveraging quantum computing, Grover's algorithm significantly speeds up the process of searching for the target sequence within the database relative to the classic methods, reducing the number of required comparisons. The execution of Qubio.py takes place on a quantum computing cloud platform.

After the algorithm runs, the results of the analysis are obtained. These results indicate the presence or location of the target sequence within the database, which is then processed to extract meaningful insights. In scientific research, understanding and analysing genetic sequences can

lead to insights into biological mechanisms, the study of genetic diseases, and the development of new therapies and bioactive molecules. In agriculture, genetic analyses can help identify the genes that are responsible for desirable plant traits, such as disease resistance, productivity, and product quality. Additionally, genetic analyses can be used in environmental monitoring cases to track biodiversity and assess the health of ecosystems by detecting and monitoring species and genetic variations.

This code implements Grover's algorithm to efficiently search for a specific target sequence within a database of DNA sequences. By leveraging Grover's transformation, which is represented mathematically by matrices, the algorithm amplifies the probability of locating the desired sequence.

Sequence definition

To define the short-length target sequence (50 bp), a protein sequence isolated from the Brazilian biomes metagenome project was explored and analysed in terms of its functional domain. Its most genetically conserved portion was highlighted and filtered to serve as a search target (Fig. 8). The chosen protein, Cas9-like nuclease, is involved in the CRISPR/Cas9 system, which has several advantages, such as its high editing efficiency, its ease of operation, its cost-effectiveness, and the diversity of its recognition sites. It can be used to edit the human genome for treating genetic defects or illnesses caused by gene mutations. It can also edit plant genes to create more resilient plants. Furthermore, the CRISPR/Cas9 editing technology offers the possibility of eliminating pathogens such as bacteria and viruses, which could completely cure infectious diseases[53].

Qubio.py code and quantum circuit

According to the reported data, Grover's algorithm[54] is the most appropriate algorithm for performing searches in a quantum environment. In the circuit of this algorithm, each qubit is input, |0> and |1> possess equal superposition probabilities, and this qubit passes, or each element is searched. The output generates the results. This superposition is achieved via the Hadamard transformation, enabling the qubit to explore multiple states simultaneously, which is a key feature for exploiting quantum parallelism. Each pass through the algorithm incrementally increases the probability of finding the correct result, enhancing the efficiency beyond that of the classic search methods. The representation of the quantum circuit for the target sequence of 50 nucleotides can be observed in Fig. S1, as along with the code used to generate it. Our code allows the search to be performed with up to a 30% mismatch rate, meaning that the similarity between the target sequence and the database sequence is 70%. In practical terms, if the query sequence and the database sequence have 100 nucleotides, a similarity of 70% indicates that 70 nucleotides are identical or very similar, whereas 30 nucleotides differ. Depending on the context of the search, a similarity of 70% may be sufficient for inferring homologies but may require an additional analysis.

Grover algorithm

The oracle implementation uses a series of control-X gates (CX gates) to label qubits by comparing the input and target sequences. The Grover diffusion operator, (U_G), amplifies the probability of obtaining the correct answer by reflecting the quantum state about the average of all the superposed states. Mathematically, the Grover transformation is represented as follows:

$$U_G = (2|s\rangle\langle s|) - I$$

where (s) represents the initial state of the uniform superposition process and (I) is the identity matrix. This transformation iterates multiple times to maximize the amplitude of the desired

quantum state. The probability (P_n) of finding the target sequence after (n) iterations of Grover's transformation is given by

$$P_n = \sin^2((2n+1)\theta)$$

where θ is the angle between the initial state $|s\rangle$ and the target state $|w\rangle$, which encodes the solution in the Hilbert space. After a sufficient number of iterations, the superposition collapses into the target state, and the binary result is decoded to retrieve the target sequence. The binary result is then converted back to the corresponding nucleotide sequence, and the result corresponding to the target sequence is checked.

Hadamard transformation

This function implements the core logic of Grover's algorithm. It begins by initializing a quantum circuit with numerous qubits (double the length of the target sequence to be searched). Hadamard gates are applied to create an initial superposition across all possible states. The oracle function is then executed to mark the target sequence. The Grover transformation is repeatedly applied to amplify the probability of measuring the correct target sequence. The final step involves measuring the quantum state and analysing the results to determine the probability of successfully identifying the target sequence.

The Grover transformation can be mathematically represented by the transformation matrix U_G, which is defined as shown below:

$$UG=2|\psi\rangle\langle\psi|-I$$

The Grover transformation, $UG=2|\psi\rangle\langle\psi|-I$, operates by reflecting the current quantum state over the average of all possible states, where $|\psi\rangle$ is the initial state of the uniform superposition process, $\langle\psi|$ is its conjugate transpose, and I is the identity matrix.

Oracle function

The oracle function applies the oracle operation to the quantum circuit and checks whether the current measurement matches the target sequence. The operation is mathematically represented by the transformation matrix U_f, which is defined as follows:

$$U_f |x\rangle |y\rangle = |x\rangle |y \oplus f(x)\rangle$$

In this equation, (x) represents the qubits encoding the input (e.g., the DNA sequence in a bioinformatics application), whereas (y) denotes the label qubit. The function f(x) returns a value of 1 if the input state (x) matches the target sequence and 0 otherwise. The oracle is implemented using CX gates, which apply the appropriate transformation by evaluating the input and target sequences. This process flags the matching state without collapsing the quantum superposition.

Measuring qubits

We implemented Grover's algorithm on real quantum hardware and compared the results with those of classic simulation schemes. The quantum simulations were performed using the Qiskit framework, leveraging the Qiskit Aer simulator for up to 100 qubits. Real quantum hardware tests were conducted on IBM Cloud's quantum systems, where access was constrained by both time (ten-minute sessions) and the number of qubits required just for the first code, Qubio.py.

Despite these constraints, the quantum implementation demonstrated the ability of the algorithm to efficiently identify target sequences with far fewer comparisons than that required by the classic algorithms. By running the simulations via Google Colab, which hosts the Qubio.py code, we highlighted the potential advantages of quantum-based algorithms in scaling up computational tasks, even in bioinformatics applications.

Scars and the BioBloQu code

To search for scar sequences and insert genetic blocks using quantum computing, the target sequences and the database sequence were defined as follows.

Two predefined sequences (left end: AAAATCTGTCATAAATTATC; right end: ATTATTCTCCTTTCTTTAGT) from a specific essential gene in a determined region of the minimal genome of *M. mycoides* JCVI-Syn3 (3906 nucleotides) were analysed. Part of the minimal cell genome was used as a database for the search. The target sequences were represented as binary vectors, where each nucleotide was mapped to a quantum state: adenine (A) corresponded to state $|1\rangle$, cytosine (C) corresponded to a superposition state, guanine (G) corresponded to a combination of $|1\rangle$ and $|0\rangle$, and thymine (T) corresponded to $|0\rangle$. The quantum circuit was constructed using the QuantumCircuit class, where Hadamard operations and controlled gates were applied to prepare the superposition and perform the search.

The insertion of the genetic block was accomplished by identifying the position of the scar in the database sequence and replacing it with the BioBloQu block. The circuit was then simulated using the Qiskit backend, where measuring the qubits resulted in counts that represented the probabilities of finding the target sequences.

All the codes were run in the IBM simulation environment via Jupyter notebooks and Google Colab.

Analysis of the results

The raw data and code generated and analysed in this study are available from the corresponding author upon request.

Additional applications

While Grover's algorithm was applied to the field of bioinformatics in our study, the quantum search algorithm can be adapted for solving broader scientific challenges. For example, it could be used to identify patterns in astrobiology datasets or search for signals in massive astronomical

data. As quantum computing technology matures, Grover's algorithm has the potential to overcome the computational limitations that were previously faced by classic systems.

Quantum searches can also enable the identification of novel traits within genetic sequences that were previously inaccessible due to computational resource constraints. This could revolutionize fields such as genomics, pharmaceuticals, agriculture, and environmental science, driving innovation forwards across a variety of disciplines. Grover's algorithm offers a transformative approach for solving computational search problems, providing a quadratic speedup over the classic methods. Quantum computing promises to revolutionize domains such as computational biology, but significant challenges remain in terms of hardware scalability and the development of domain-specific quantum algorithms. Collaborative efforts between quantum physicists and computational scientists will be essential for overcoming these challenges, pushing quantum technology beyond the constraints of classic systems.

Data availability

All the data presented in this paper resulted from numerical simulations generated by computer programming codes. Correspondence and requests for materials should be addressed to Elibio Rech (e-mail address: elibio.rech@embrapa.br).

Code availability

The code underlying this study is not publicly available but may be made available to qualified researchers upon reasonable request from the corresponding author.

Acknowledgements

We acknowledge the National Institute of Science and Technology in Synthetic Biology, National Institute of Science and Technology in Engineering Biological Systems and the Ministry of Agriculture and Livestock. Funding from the National Council for Scientific and Technological

657 Development (465603/2014-9; 400145/2023-5), Research Support Foundation of the Federal
658 District (0193.001.262/2017), and the Coordination for the Improvement of Higher Education
659 Personnel.

660 **Competing interests**

661 The authors have no conflicts of interest to declare.