Efficient DNA coding algorithm for PCR amplification information retrieval Supplementary Materials

1.Analysis of Nonspecific Pairing constraint

In PCR amplification, thevalue represents the temperature of DNA double-strand separation, and the higher thevalue, the more stable the base binding and the less easy to separate. value refers to the free energy required for DNA double strand formation, which reflects the relative stability of base pairs inside the double strand structure. The larger the absolute value of, the more stable the double strand is[1,2].

This paper designed a large-scale simulation experiment and selectedand (in this experiment is absolute value) to explore the stability of binding between DNA sequence and primer when nonspecific pairing of different bits occurs between DNA sequence and primer 3 'end. In this paper, 880 primer sequences are used to design DNA sequences in turn, so that the DNA sequence and the 3 'end of the primer sequence at different positions produce several consecutive bits of non-specific pairing. The DNA sequence is 200bp, and the primer length is 20bp, and the averageandare calculated, as shown in **Figure 1**.

非特异性配对位数

**Figure 1.** andvalues of DNA sequences and primers for different number of nonspecific pairing bits.

The experimental results showed that the more bits of DNA sequence and primer 3 'end produced nonspecific pairing, the higherandvalues, which represented a more stable binding to the primer and conformed to the PCR amplification law. Theandvalues of DNA sequences and primers were lower when the number of unspecific pairing bits was less than 8 at the 3 'end of DNA sequences and primers, indicating that DNA sequences and primers were not easy to be unspecific pairing at this time, that is, PCR amplification was not error-prone. When the DNA sequence and the primer 3 'end produced a continuous 8 bits of non-specific pairing, theand values will rise sharply, more than 50% of the maximum value, indicating that the primer is more prone to non-specific pairing with the DNA sequence to amplify the interference sequence. When the number of non-specific pairing bits between DNA sequence and primer 3 'end exceeds 8, the combination of DNA sequence and primer is more stable and prone to the risk of amplification interference.

In order to verify the simulation results, this paper randomly selected a primer from the primer library for biochemical experiments. Ten DNA sequences with a length of 200bp were generated, and the non-specific paired base segments ofconsecutive positions () existed between the sequence and the 3 'end of the primer in the sequence and the posterior segment, as shown in **Figure 2** below.

DNA序列与引物3′端产生连续N位非特异性配对的示意图

**Figure 2.** Schematic representation of a DNA sequence and the 3 'end of a primer producing a nonspecific pairing of consecutivepositions.

The above DNA sequences were amplified by PCR in turn, and the amplification of DNA sequences under different non-specific pairing digits was analyzed. The experimental results are shown in **Table 1** below. From the experimental results, it is easy to see that as long as there is nonspecific pairing between the DNA sequence and the 3 'end of the primer, the primer will bind to the DNA sequence at the nonspecific pairing position, thus amplifying invalid short sequences. When the number of unspecific pairing bits was less than 8, the proportion of invalid short sequences was relatively small, ranging from 1.2% to 9.9%, and the success rate of PCR amplification was from 90.1% to 98.8%. However, when the number of nonspecific pairing bits reached 8 or more, the proportion of invalid short sequences increased, and the success rate of PCR amplification significantly decreased to 80.9% or less. It is proved that the non-specific pairing of DNA sequence and primer will produce invalid interference amplification sequences, resulting in a decrease in the success rate of PCR amplification. When the number of non-specific pairing bits is 8 or more, it will have a more serious impact. This experimental result further proves the influence of nonspecific pairing situation in DNA sequence on PCR amplification.

**Table 1.** Proportion of each amplified sequence in different cases of non-specific pairing number.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Amplified Sequence Type** | **Percentage(%)** | | | | | | | | | |
| =2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 图片11 | 98.8 | 97.9 | 95.2 | 93.7 | 91.5 | 90.1 | 80.9 | 77.1 | 72.7 | 69.8 |
| 图片12 | 1.2 | 2.1 | 4.3 | 4.5 | 6.2 | 7.0 | 10.7 | 12.8 | 14.9 | 16.4 |
| 图片13 | 0 | 0 | 0.5 | 1.8 | 2.3 | 2.9 | 8.4 | 10.1 | 12.4 | 13.8 |

In summary, in order to ensure the specificity of the PCR amplification reaction and improve the efficiency of PCR amplification, the generated DNA sequence should try to avoid the non-specific pairing of consecutive 8 bits with the 3 'end of the primer (hereinafter collectively referred to as "non-specific pairing constraint"). Based on this constraint, this paper proposes a novel efficient DNA coding algorithm for information retrieval of PCR amplification.

References

1. Karami, A.; Gill, P.; Kalantar Motamedi, M.; Saghafinia, M. A review of the current isothermal amplification techniques: applications, advantages and disadvantages. *Journal of global infectious diseases* **2011**, *3*, 293-302.

2. Mann, T.; Humbert, R.; Dorschner, M.; Stamatoyannopoulos, J.; Noble, W.S. A thermodynamic approach to PCR primer design. *Nucleic acids research* **2009**, *37*, e95-e95.