Multidimensional Data Organization and Random Access in Large-Scale DNA Storage Systems

Xin Song1,2,3, Shalin Shah1,3, John Reif1,3
1Department of Electrical and Computer Engineering, 2Biomedical Engineering, 3Computer Science
Duke University, Durham, NC 27708, USA

Introduction

• DNA archival storage – density, capacity, durability, data segments encoded as oligos with addresses
• PCR random access – limited by orthogonal primers, unable/inefficient to explore different data retrieval patterns
• Hierarchical memory – scalability of nested primer hierarchy, specificity of single entry retrieval

Highlights

• Combine nested and semi-nested PCR to virtually enforce multidimensional data organization in large DNA storage systems
• Efficient use/reuse of PCR primers to establish large address space (k + n primers to index nk entries)
• Strategic use of primers from different layers to specify data retrieval in form of rows, columns, tables, and blocks
• Support multiple well-defined data retrieval patterns tailored to data relations/structures
• Enrich and extract specific data entry/subset via just one or two rounds of PCR

Architecture and Operations

Number of orthogonal PCR primers needed (k + n)

= (# of rows in a table) + (# of columns in a table) + (# of tables in a block) + (# of blocks) + 2

Number of unique data entries addressable (n^2)

= (# of rows in a table) × (# of columns in a table) × (# of tables in a block) × (# of blocks)

Design Variation

Number of orthogonal PCR primers needed (3 + n)

= (# of rows in a table) + (# of columns in a table) + (# of tables ×2) + 2

Number of unique data entries addressable (n^2)

= (# of rows in a table) × (# of columns in a table) × (# of tables)

Highlights

- Combine nested and semi-nested PCR to virtually enforce multidimensional data organization in large DNA storage systems
- Efficient use/reuse of PCR primers to establish large address space (k + n primers to index nk entries)
- Strategic use of primers from different layers to specify data retrieval in form of rows, columns, tables, and blocks
- Support multiple well-defined data retrieval patterns tailored to data relations/structures
- Enrich and extract specific data entry/subset via just one or two rounds of PCR

This work was sponsored by NSF grant no. CCF 1617791. X.S. acknowledges support from NSF grant no. DGE 1545220. We would like to thank Xiangyu Zhang and Dr. Joshua Granek for helpful discussions.