

УДК 575.112:004

**Анна Ляшенко, Олександр Богатирьов**

Черкаський національний університет імені Богдана Хмельницького, Україна

## **ВИКОРИСТАННЯ АЛГОРИТМУ БАРОУЗА-УІЛЕРА ДЛЯ ВИРІВНЮВАННЯ ГЕНЕТИЧНИХ ПОСЛІДОВНОСТЕЙ**

**Anna Liashenko, Oleksandr Bogatyrev**

### **BURROWS-WHEELER ALGORITHM FOR ALIGNING THE GENETIC SEQUENCES**

Вирівнюванням послідовностей називають аналіз подібностей та відмінностей на рівні окремих основ та амінокислот, який приводять з метою встановлення структурних, функціональних та еволюційних відношень між генетичними послідовностями. Інакше кажучи, вирівнювання є визначення взаємної відповідності залишків. При цьому, будь-яка встановлена відповідність, при якій зберігається вихідний порядок залишків в послідовності, є їх вирівнюванням. Якщо послідовності мають деяку спільну предкову послідовність, то вони, як правило, виявляють подібність в поєднаннях мономерів, структурах та біологічних функціях. Наприклад, якщо відкрито нову послідовність з невідомою функцією і при цьому в базах даних можуть бути знайдені подібні їй послідовності з раніше встановленими структурами та функціями, то вирівнювання може стати основою для передбачування функції або структури цієї послідовності, що є особливо актуальним для задач біоінформатики.

Нові технології, які використовуються для вирівнювання ДНК, генерують надзвичайно велику кількість коротких послідовностей. При цьому виникає потреба в швидких програмах для вирівнювання кожної з розглянутих послідовностей. Перше покоління методів, побудованих на хеш-таблицях, включаючи MAQ, є достатньо швидкими для процедури вирівнювання. Тим не менше, MAQ не підтримує розривне вирівнювання для записів з одним закінченням, що робить цей метод непридатним для вирівнювання довгих послідовностей, в яких часто виникають вставки та видалення. Як альтернатива методу MAQ в даній роботі розглядається новий спосіб вирівнювання послідовностей, який базується на перетворенні Барроуза-Уілера (Burrows–Wheeler Transform - BWT). Аналіз показує, що хоча теоретично алгоритм вирівнювання BWA (Burrows–Wheeler Aligner) працює з довільно довгими послідовностями, його продуктивність на практиці знижується на довгих послідовностях, особливо коли значення помилки в послідовності велике. Крім того, BWA завжди вимагає, щоб вирівнювалась повна послідовність, з першої основи і до останньої, але довші послідовності більш ймовірно можуть бути порушені структурними змінами в геномі, які не можуть бути обраховані BWA. В результаті розрахунків показано, що кращим виходом для аналізу довгих послідовностей буде розділити послідовність на значну кількість коротких фрагментів, вирівняти фрагменти окремо у відповідності до описаного алгоритму та приєднати часткове вирівнювання для отримання повного вирівнювання послідовності.

Пошук подібностей в базах даних дає нам можливість отримати послідовності, які є подібними до послідовності запиту. Отримана послідовність, для якої відома функціональна і структурна інформація, допоможе нам передбачити структуру і функцію послідовності запиту. При звичайному перегляді бази даних послідовність запиту вирівнюється з кожною послідовністю бази даних. В загальному випадку перегляд бази даних необхідно проводити з метою відшукування гомологій.