

# Concorrenca e Paralelismo. Bloque II Paralelismo

## Práctica 3: Descomposición de dominio: vector de similaridad de conjuntos de ADN

Primavera 2023



## Vector de similitud de conjuntos de ADN

### Datasets de secuencias de ADN

- M secuencias de texto
- Cada una con N bases (caracteres A, C, G, T, N)

### Vector de similitud

- Datos dos conjuntos de secuencias de ADN
- Medida de lo similares que son la secuencia  $i$  del primer conjunto y la secuencia  $j$  del segundo conjunto
  - M resultados: Los dos conjuntos deben tener el mismo número M de secuencias
  - SIMPLIFICACION: Todas las secuencias tienen el mismo número N de bases

# Vector de similitud de conjuntos de ADN

## Código secuencial

```
int main(int argc, char *argv[]){
    int *data1, *data2, *result;
    data1 = (int *) malloc(M*N*sizeof(int));
    data2 = (int *) malloc(M*N*sizeof(int));
    result = (int *) malloc(M*sizeof(int));

    /* Initialize Matrices */
    ...

    for(i=0;i<M;i++) {
        result[i]=0;
        for(j=0;j<N;j++) {
            result[i] += base_distance(data1[i*N+j], data2[i*N+j]);
        }
    }

    free(data1); free(data2); free(result);
}
```

# Vector de similitud de conjuntos de ADN

Código secuencial: `similarity.c`

## Descomposición de dominio

- Dividir las matrices entre  $p$  procesos, cada uno con  $rows = M/p$  filas (por simplicidad, empezad con el caso en que consideraremos que el número de procesos es múltiplo de  $M$ :  $M \bmod p = 0$ ).
- Cada tarea se encargará de calcular  $M/p$  elementos del vector resultado.
- En la entrega considerad cualquier número de procesos, no solo múltiplos de  $M$ .