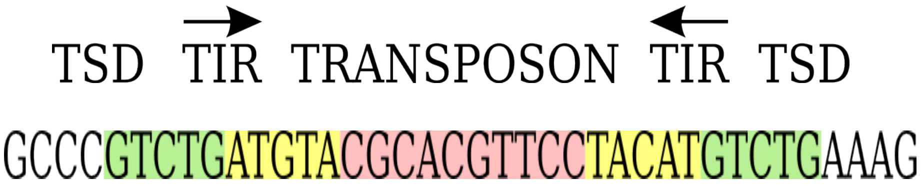


TSD TIR TRANSPOSON TIR TSD

GCCCGTCTGATGTACGCACGTTCCTACATGTCTGAAAG

```
key_locs = locate_tir_keys(sequence)
suffix_tree_create(sequence)
for all key_locs do
    ic = calculate_inverted_complement()
    ic_locs = suffix_tree_find(ic)
    for all ic_locs do
        if ( tsd_match(key_loc, ic_loc) ) then
            store_location(key_loc, ic_loc)
        end if
    end for
end for
report_locations_to_user()
```

Trolling for Transposons

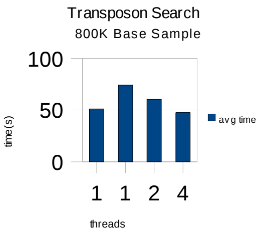
Derrick Kearney

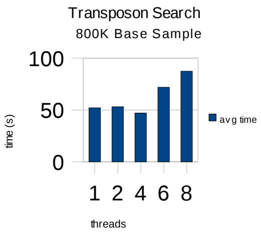
dsk@purdue.edu

Purdue University

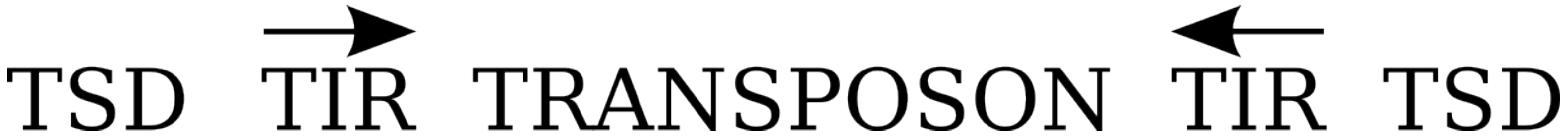
ECE 563

April 16, 2009





Identifying Transposons



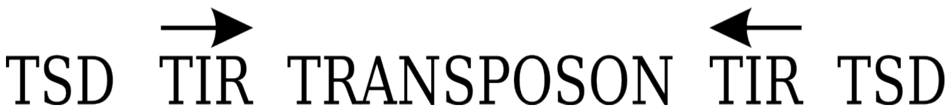
TSD TIR TRANSPOSON TIR TSD

TSD - Target Site Duplication
TIR - Terminal Inverted Repeat

A ↔ T
C ↔ G

GCCCGTCTGATGTACGCACGTTCCTACATGTCTGAAAG

Approach



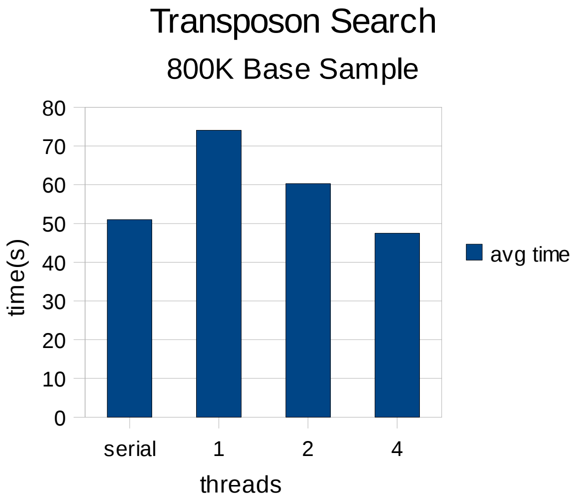
TSD TIR TRANSPOSON TIR TSD

```
key_locs = locate_tir_keys(sequence)
suffix_tree_create(sequence)
for all key_locs do
    ic = calculate_inverted_complement()
    ic_locs = suffix_tree_find(ic)
    for all ic_locs do
        if ( tsd_match(key_loc, ic_loc) ) then
            store_location(key_loc, ic_loc)
        end if
    end for
end for
report_locations_to_user()
```

Test Setup

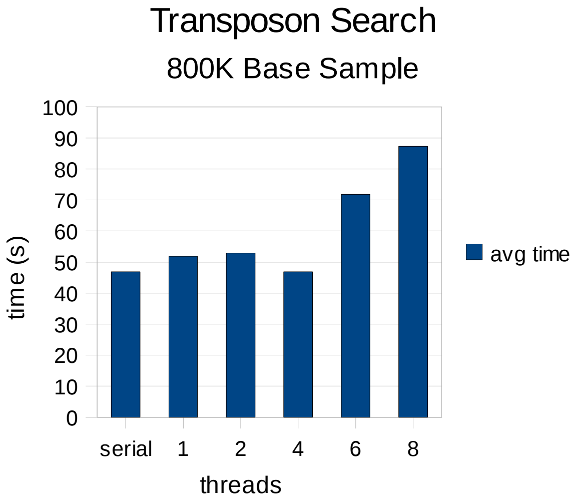
Machine A:

- 2 x Dual-Core AMD Opteron
  - Processor Model 2222
  - (3GHz / 2 x 1MB – L2)



Machine B:

- 2 x Quad-Core Intel Xeon
  - Processor X5355
  - (2.66Ghz / 2 x 4MB - L2)



Future Work

- Lower memory footprint
- New ways to report uniques

GCCCGTCTGATGTACGCACGTTCCTACATGTCTGAAAG

GCCCGTCTGATGTACGCACGTTCCTACATGTCTGAAAG

GCCCGTCTGATGTACGCACGTTCCTACATGTCTGAAAG