

Report fsm-lite

Profa Dra Helena R. S. D’Espindula

2025-08-04

Contents

1	Considerações Gerais	3
2	Programa original	3
2.1	Arquivo fsm-lite.cpp	3
2.2	Arquivo Makefile	7
3	Versão 1.0 (original)	8
3.1	Objetivo nessa versão:	8
3.2	Pre-requisitos:	8
3.3	Compilação do fsm-lite e Teste de funcionamento basico	11
3.4	Criacao de controle de versao no github (fork do original):	17
3.5	Criação de arquivo .sh para excussão padronizada	17
3.6	Avaliação dos logs de output	21
4	Versão 2.0	32
4.1	Modificações	32
4.2	Comparação dos arquivos de resultado	33
4.3	Avaliação dos logs de output (comparação de desempenho)	33

```
knitr::opts_chunk$set(echo = TRUE)
require(tidyverse)
```

```
## Carregando pacotes exigidos: tidyverse
```

```
## Warning: pacote 'ggplot2' foi compilado no R versão 4.4.3
```

```
## Warning: pacote 'tibble' foi compilado no R versão 4.4.3
```

```
## Warning: pacote 'purrr' foi compilado no R versão 4.4.3
```

```
## Warning: pacote 'lubridate' foi compilado no R versão 4.4.2
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.2      v tibble    3.3.0
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.1.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
require(stringr)
require(data.table)
```

```
## Carregando pacotes exigidos: data.table
```

```
## Warning: pacote 'data.table' foi compilado no R versão 4.4.3
```

```
##
## Anexando pacote: 'data.table'
##
## Os seguintes objetos são mascarados por 'package:lubridate':
##
##   hour, isoweek, mday, minute, month, quarter, second, wday, week,
##   yday, year
##
## Os seguintes objetos são mascarados por 'package:dplyr':
##
##   between, first, last
##
## O seguinte objeto é mascarado por 'package:purrr':
##
##   transpose
```

```
require(lubridate)
require(knitr)
```

```
## Carregando pacotes exigidos: knitr
```

```
## Warning: pacote 'knitr' foi compilado no R versão 4.4.3
```

```
require(kableExtra)
```

```
## Carregando pacotes exigidos: kableExtra
```

```
## Warning: pacote 'kableExtra' foi compilado no R versão 4.4.2
```

```
##
## Anexando pacote: 'kableExtra'
##
## O seguinte objeto é mascarado por 'package:dplyr':
##
##      group_rows

require(formatR)

## Carregando pacotes exigidos: formatR

## Warning: pacote 'formatR' foi compilado no R versão 4.4.3

require(tinytex)

## Carregando pacotes exigidos: tinytex

## Warning: pacote 'tinytex' foi compilado no R versão 4.4.3

require(ggplot2)
require(dplyr)
```

1 Considerações Gerais

Os arquivos do fsm-lite originais podem ser encontrados no Github: <https://github.com/nvalimak/fsm-lite>
 Para facilitar o entendimento apenas os arquivos originais de maior importancia foram transcritos no item a seguir.

2 Programa original

2.1 Arquivo fsm-lite.cpp

Arquivo: original/fsm-lite.cpp

```
#include "default.h"
#include "configuration.h"
#include "input_reader.h"
#include <sdsl/suffix_trees.hpp> // TODO: replace with csa+lcp array
#include <sdsl/wt_algorithm.hpp>
#include <iostream>
#include <vector>
#include <cstdlib> // std::exit()
using namespace std;

typedef sdsl::cst_sct3<> cst_t;
typedef sdsl::wt_int<> wt_t;
typedef sdsl::bit_vector bitv_t;
```

```

typedef cst_t::char_type char_type;
typedef cst_t::node_type node_type;
typedef wt_t::size_type size_type;

/**
 * Construct the sequence labels
 *
 * Assumes that the number of input files is less than 2DBITS.
 * The value of DBITS has to be set at compile time (in defaults.h).
 * Large DBITS values result in large memory requirements for wt_init().
 */
void wt_init(wt_t &wt, bitv_t &separator, cst_t &cst, input_reader *ir, configuration &config)
{
    uint64_t n = cst.csa.size();
    sds1::int_vector<DBITS> labels(n, ~0u);
    separator = bitv_t(n, 0);
    uint64_t k = ir->size()-1;
    uint64_t j = cst.csa.wavelet_tree.select(1, 0);
    if (config.debug)
        cerr << "bwt end marker pos = " << j << endl;
    uint64_t bwtendpos = j;
    j = cst.csa.lf[j];
    labels[j] = 0; // Label of last byte
    separator[n-1] = 0;
    separator[n-2] = 1;
    j = cst.csa.lf[j];
    for (uint64_t i = n-2; i > 0; i--) {
        char_type c = cst.csa.bwt[j];
        labels[j] = k;
        if (c == '$')
            k--;
        if (c == '$' || c == '#')
            separator[i-1] = 1;

        j = cst.csa.lf[j];
    }
    labels[j] = k;
    if (j != bwtendpos || k != 0) // Assert
    {
        cerr << "Labeling failed, j = " << j << ", k = " << k << endl;
        exit(1);
    }

    //TODO cleanup
    /*for (uint64_t i = 0; i < n; ++i)
        cerr << cst.csa.text[i];
    cerr << endl;
    for (uint64_t i = 0; i < n; ++i)
        cerr << separator[i];
    cerr << endl;
    for (uint64_t i = 0; i < n; ++i)
        cerr << labels[cst.csa.isa[i]];
    cerr << endl;
*/
}

```

```

*/

std::string tmp_file = sds1::ram_file_name(sds1::util::to_string(sds1::util::pid())+"_"+sds1::util::
sds1::store_to_file(labels, tmp_file);
sds1::int_vector_buffer<DBITS> text_buf(tmp_file);
wt = wt_t(text_buf, labels.size());
if (config.debug)
    cerr << "wt size = " << wt.size() << ", n = " << n << endl;
j = 0;
for (uint64_t i = 0; i < ir->size(); ++i)
    j += wt.rank(n, i);
if (j != n) // Assert
{
    cerr << "Label sum failed, j = " << j << ", n = " << n << endl;
    exit(1);
}
}

int main(int argc, char ** argv)
{
    configuration config(argc, argv);
    if (!config.good)
        config.print_short_usage();

    if (config.verbose)
        cerr << "Reading input files..." << endl;
    input_reader *ir = input_reader::build(config);
    if (config.verbose)
        cerr << "Read " << ir->size() << " input files and " << ir->total_seqs() << " sequences of total

/**
 * Initialize the data structures
 */
if (config.verbose)
    cerr << "Constructing the data structures..." << endl;
cst_t cst;
construct(cst, config.tmpfile + ".tmp", 1);
if (!cst.csa.size())
{
    cerr << "error: unable to construct the data structure; out of memory?" << endl;
    abort();
}

wt_t label_wt;
bitv_t separator;
wt_init(label_wt, separator, cst, ir, config);

bitv_t::rank_1_type sep_rank1(&separator);
//bitv_t::select_1_type sep_select1(&separator); TODO Remove?
assert(sep_rank1(cst.size()) == ir->total_seqs());

size_type support = 0;

```

```

vector<wt_t::value_type> labels(ir->size(), 0);
vector<size_type> rank_sp(ir->size(), 0);
vector<size_type> rank_ep(ir->size(), 0);

if (config.verbose)
    cerr << "Construction complete, the main index requires " << size_in_mega_bytes(cst) << " MiB p

/**
 * Main loop
 */
node_type root = cst.root();
vector<node_type> buffer;
buffer.reserve(1024*1024);
for (auto& child: cst.children(root))
    buffer.push_back(child);
while (!buffer.empty())
{
    node_type const node = buffer.back();
    buffer.pop_back();
    unsigned depth = cst.depth(node);
    if (depth < config.maxlength)
        for (auto& child: cst.children(node))
            buffer.push_back(child);
    if (depth < config.minlength)
        continue;
    if (cst.is_leaf(node))
        continue;

    // Process the candidate node
    size_type sp = cst.lb(node);
    size_type ep = cst.rb(node);
    node_type wn = cst.wl(node, cst.csa.bwt[sp]);
    /*if (config.debug)
    {
        size_type pos = cst.csa[sp];
        auto s = extract(cst.csa, pos, pos + depth - 1);
        cerr << "at node = " << depth << "-[" << sp << "," << ep << "], wl = " << (wn != root);
        if (wn!=root)
            cerr << "[" << cst.rb(wn)-cst.lb(wn) << " vs " << ep-sp << "];
        cerr << ", seq = " << s << endl;
    }*/
    if (wn == root && config.debug)
    {
        cerr << "warning: no Weiner-link at " << depth << "-[" << sp << "," << ep << "]" << endl;
        continue;
    }
    if (depth < config.maxlength && cst.rb(wn)-cst.lb(wn) == ep-sp)
        continue; // not left-branching

    sds1::interval_symbols(label_wt, sp, ep+1, support, labels, rank_sp, rank_ep);
    if (support < config.minsupport || support > config.maxsupport)
        continue;

```

```

size_type truesupp = 0;
for (size_type i = 0; i < support; ++i)
    if (config.minfreq <= rank_ep[i]-rank_sp[i])
        ++truesupp;
if (truesupp < config.minsupport)
    continue;

if (depth > config.maxlength)
    depth = config.maxlength;
size_type pos = cst.csa[sp];
// Check for separator symbol TODO cleanup
/*unsigned p_depth = cst.depth(cst.parent(node));
if (sep_rank1(pos) != sep_rank1(pos + p_depth))
    continue; // Separator occurs above parent node
if (sep_rank1(pos) != sep_rank1(pos + depth))
    depth = sep_select1(sep_rank1(pos)+1) - pos + 1; // Separator above current node
*/

if (sep_rank1(pos) != sep_rank1(pos + depth))
    continue;
auto s = extract(cst.csa, pos, pos + depth - 1);
if (input_reader::smaller_than_rev_cmpl(s))
    continue;
cout << s + " |";
for (size_type i = 0; i < support; ++i)
    if (config.minfreq <= rank_ep[i]-rank_sp[i])
        cout << ' ' << ir->id(labels[i]) << ':' << rank_ep[i]-rank_sp[i];
cout << '\n';
}

if (config.verbose)
    cerr << "All done." << endl;
delete ir; ir = 0;
return 0;
}

```

2.2 Arquivo Makefile

Arquivo: original/Makefile

```

SDSL_INSTALL_PREFIX=${HOME}/software

CPPFLAGS=-std=c++11 -I$(SDSL_INSTALL_PREFIX)/include -DNDEBUG -O3 -msse4.2
LIBS=-ldsl -ldivsufsort -ldivsufsort64
OBJ = configuration.o input_reader.o fsm-lite.o

fsm-lite: $(OBJ)
    $(LINK.cpp) $^ -L$(SDSL_INSTALL_PREFIX)/lib $(LIBS) -o $@

test: fsm-lite
    ./fsm-lite -l test.list -t tmp -v --debug -m 1

clean:

```

```
$(RM) fsm-lite *.o *~

depend:
    g++ -MM -std=c++11 -I$(SDSL_INSTALL_PREFIX)/include *.cpp > dependencies.mk

include dependencies.mk
```

3 Versão 1.0 (original)

3.1 Objetivo nessa versão:

- Instalação de pre-requisitos
- Compilação
- Teste de funcionamento basico
- Controle de versão em Github
- Script de excussão com monitoramento padronizado
- Testes com diferentes Ns e parametros

3.2 Pre-requisitos:

- Ter listagem de genomas
- Ter acesso e permissão de leitura dos genomas
- Ter o sds1-lite v2.0.3 (versão recomendada pelo fsm-lite original) instalado na home

Fonte do sds1-lite v2.0.3: <https://github.com/simongog/sdsl-lite/releases/tag/v2.0.3>

Instalação do sds1-lite v2.0.3:

```
helena.despindula@BIOINF008:~$ cd ~/sdsl-lite-2.0.3

helena.despindula@BIOINF008:~$ mkdir -p build

helena.despindula@BIOINF008:~/sdsl-lite-2.0.3$ cd build

helena.despindula@BIOINF008:~/sdsl-lite-2.0.3/build$ cmake .. -DCMAKE_INSTALL_PREFIX=$HOME/sdsl-lite-2.0.3
-- Compiler is recent enough to support C++11.
-- Performing Test HAVE_GCC_STD=C__11_WALL_WEXTRA__DNDEBUG
-- Performing Test HAVE_GCC_STD=C__11_WALL_WEXTRA__DNDEBUG - Success
CMake Warning (dev) at external/gtest-1.6.0/CMakeLists.txt:42 (project):
  Policy CMP0048 is not set: project() command manages VERSION variables.
  Run "cmake --help-policy CMP0048" for policy details.  Use the cmake_policy
  command to set the policy and suppress this warning.

The following variable(s) would be set to empty:

  PROJECT_VERSION
  PROJECT_VERSION_MAJOR
  PROJECT_VERSION_MINOR
  PROJECT_VERSION_PATCH
This warning is for project developers.  Use -Wno-dev to suppress it.
```



```
CMake Warning (dev) at external/libdivsufsort-2.0.1/CMakeLists.txt:19 (project):
Policy CMP0048 is not set: project() command manages VERSION variables.
Run "cmake --help-policy CMP0048" for policy details. Use the cmake_policy
command to set the policy and suppress this warning.
```

The following variable(s) would be set to empty:

```
PROJECT_VERSION
PROJECT_VERSION_MAJOR
PROJECT_VERSION_MINOR
PROJECT_VERSION_PATCH
```

This warning is for project developers. Use `-Wno-dev` to suppress it.

```
-- Configuring done
-- Generating done
-- Build files have been written to: /home/helena.despindula/sdsl-lite-2.0.3/build
```

```
helena.despindula@BI0INFO08:~/sdsl-lite-2.0.3/build$ make -j$(nproc)
```

```
[ 4%] Built target gtest
[ 15%] Built target divsufsort64
[ 27%] Built target divsufsort
[ 95%] Built target sdsl
[100%] Built target gtest_main
```

```
helena.despindula@BI0INFO08:~/sdsl-lite-2.0.3/build$ make install
```

```
[ 4%] Built target gtest
[ 9%] Built target gtest_main
[ 20%] Built target divsufsort64
[ 31%] Built target divsufsort
[100%] Built target sdsl
```

Install the project...

```
-- Install configuration: "Release"
```

```
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/divsufsort.h
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/divsufsort64.h
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/lib/libdivsufsort.a
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/lib/libdivsufsort64.a
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/bit_vector_il.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/bit_vectors.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/bits.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/bp_support.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/bp_support_algorithm.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/bp_support_g.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/bp_support_gg.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/bp_support_sada.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/coder.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/coder_comma.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/coder_elias_delta.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/coder_elias_gamma.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/coder_fibonacci.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/config.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/construct.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/construct_bwt.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/construct_config.hpp
```

```

-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/construct_isa.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/construct_lcp.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/construct_lcp_helper.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/construct_sa.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/construct_sa_se.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/csa_alphabet_strategy.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/csa_bitcompressed.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/csa_sada.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/csa_sampling_strategy.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/csa_wt.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/cst_iterators.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/cst_sada.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/cst_sct3.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/dac_vector.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/enc_vector.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/fast_cache.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/int_vector.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/int_vector_buffer.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/int_vector_io_wrappers.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/int_vector_mapper.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/inv_perm_support.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/io.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/iterators.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/k2_treap.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/k2_treap_algorithm.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/k2_treap_helper.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/lcp.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/lcp_bitcompressed.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/lcp_byte.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/lcp_dac.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/lcp_support_sada.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/lcp_support_tree.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/lcp_support_tree2.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/lcp_vlc.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/lcp_wt.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/loads_tree.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/memory_management.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/nearest_neighbour_dictionary.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/nn_dict_dynamic.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/qsufsort.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/ram_filebuf.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/ram_fs.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/rank_support.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/rank_support_scan.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/rank_support_v.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/rank_support_v5.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/rmq_succinct_sada.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/rmq_succinct_sct.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/rmq_support.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/rmq_support_sparse_table.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/rrr_helper.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/rrr_vector.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/rrr_vector_15.hpp

```

```

-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/sd_vector.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/sdsl_concepts.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/select_support.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/select_support_mcl.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/select_support_scan.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/sfstream.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/sorted_int_stack.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/sorted_multi_stack_support.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/sorted_stack_support.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/structure_tree.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/suffix_array_algorithm.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/suffix_array_helper.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/suffix_arrays.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/suffix_tree_algorithm.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/suffix_tree_helper.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/suffix_trees.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/uint128_t.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/uint256_t.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/uintx_t.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/util.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/vectors.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/vlc_vector.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/wavelet_trees.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/wm_int.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/wt_algorithm.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/wt_bldc.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/wt_gmr.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/wt_helper.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/wt_huff.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/wt_hutu.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/wt_int.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/wt_pc.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/include/sdsl/wt_rlmn.hpp
-- Installing: /home/helena.despindula/sdsl-lite-2.0.3/lib/libsdsl.a

helena.despindula@BI0INFO08:~/sdsl-lite-2.0.3/build$ ls ~/sdsl-lite-2.0.3/lib/lib*.a
/home/helena.despindula/sdsl-lite-2.0.3/lib/libdivsufsort64.a  /home/helena.despindula/sdsl-lite-2.0.3/

helena.despindula@BI0INFO08:~/sdsl-lite-2.0.3/build$ cd

helena.despindula@BI0INFO08:~$ cd LACTAS-HELISSON-01/Helena-stuff/fsm-lite/v2-0/

helena.despindula@BI0INFO08:~/LACTAS-HELISSON-01/Helena-stuff/fsm-lite/v2-0$ ls
configuration.cpp  default.h          execussao_padronizada_v2_0.sh  input_fsm-lite_OXA-23_OXA-2
configuration.h    dependencies.mk    fsm-lite.cpp                  input_fsm-lite_OXA-23_OXA-2
configuration.o    execussao_padronizada.sh  fsm-lite.o                   input_reader.cpp

```

3.3 Compilação do fsm-lite e Teste de funcionamento basico

3.3.1 Github

O fsm-lite foi primeiramente obtido do github e tentado instalar com make depend && make conforme instruções do README.mk

- Ficou assim:

Execução do **make** no terminal com compilação bem sucedida:

Então tentou-se uma excussão simples por linha de comando direta para teste.

```
helena.despindula@BIOINFO08:~/LACTAS-HELISSON-01/Helena-stuff/fsm-lite/v2-0$ time ./fsm-lite -l input_files
Reading input files...
Read 10 input files and 1154 sequences of total length 80325904 (includes rev.compl. sequences)
Constructing the data structures...
Construction complete, the main index requires 230.716 MiB plus 56.9204 MiB for labels.
All done.
```

```
real    0m59,317s
user    0m55,163s
sys     0m3,206s
```

```
helena.despindula@BIOINFO08:~/LACTAS-HELISSON-01/Helena-stuff/fsm-lite/v2-0$ ls -a -l
total 80032
```

```
drwxrwxr-x 2 helena.despindula helena.despindula 8192 jul 23 22:19 .
drwxrwxr-x 7 helena.despindula helena.despindula 8192 jul 23 22:01 ..
-rw-rw-r-- 1 helena.despindula helena.despindula 4894 abr 12 2016 configuration.cpp
-rw-rw-r-- 1 helena.despindula helena.despindula 557 jul 22 22:44 configuration.h
-rw-rw-r-- 1 helena.despindula helena.despindula 20712 jul 23 22:15 configuration.o
-rw-rw-r-- 1 helena.despindula helena.despindula 173 abr 12 2016 default.h
-rw-rw-r-- 1 helena.despindula helena.despindula 7639 jul 23 22:15 dependencies.mk
-rwxrwxr-x 1 helena.despindula helena.despindula 2714 jul 22 14:32 execussao_padronizada_v2_0_2.sh
-rwxrwxr-x 1 helena.despindula helena.despindula 3289 jul 23 22:01 execussao_padronizada_v2_0_3.sh
-rwxrwxr-x 1 helena.despindula helena.despindula 3294 jul 23 22:01 execussao_padronizada_v2_0_4.sh
-rwxrwxr-x 1 helena.despindula helena.despindula 2571 jul 22 14:29 execussao_padronizada_v2_0.sh
-rwxrwxr-x 1 helena.despindula helena.despindula 778176 jul 23 22:16 fsm-lite
-rw-rw-r-- 1 helena.despindula helena.despindula 6468 jul 23 22:14 fsm-lite.cpp
-rw-rw-r-- 1 helena.despindula helena.despindula 482424 jul 23 22:16 fsm-lite.o
-rw-rw-r-- 1 helena.despindula helena.despindula 1449 jul 22 15:39 input_fsm-lite_OXA-23_OXA-24_010
-rw-rw-r-- 1 helena.despindula helena.despindula 2888 jul 22 16:03 input_fsm-lite_OXA-23_OXA-24_020
-rw-rw-r-- 1 helena.despindula helena.despindula 5669 abr 12 2016 input_reader.cpp
-rw-rw-r-- 1 helena.despindula helena.despindula 1676 jul 22 22:44 input_reader.h
-rw-rw-r-- 1 helena.despindula helena.despindula 38632 jul 23 22:15 input_reader.o
-rw-rw-r-- 1 helena.despindula helena.despindula 35141 abr 12 2016 LICENSE.md
-rw-rw-r-- 1 helena.despindula helena.despindula 613 jul 22 13:30 Makefile
-rw-rw-r-- 1 helena.despindula helena.despindula 1301 abr 12 2016 README.md
-rw-rw-r-- 1 helena.despindula helena.despindula 250 jul 23 22:16 temp.meta
-rw-rw-r-- 1 helena.despindula helena.despindula 80325904 jul 23 22:16 temp.tmp
```

3.3.2 Conda

Devido aos problemas de instalação tentamos usar a versão do conda do sds1-lite e do fsm-lite no ambiente fsm-lite-conda.

Instalação:

```
(fsm-lite-conda) helena.despindula@BIOINFO08:~$ conda install conda-forge::sds1-lite
/usr/local/miniconda3/lib/python3.9/site-packages/conda/base/context.py:211: FutureWarning: Adding 'def
```

To remove this warning, please choose a default channel explicitly with conda's [regular configuration system](#).

```
conda config --add channels defaults
```

For more information see <https://docs.conda.io/projects/conda/en/stable/user-guide/configuration/use-com>

```
deprecatad.topic(
Channels:
- defaults
- conda-forge
Platform: linux-64
```

```
Collecting package metadata (repodata.json): done
Solving environment: done
```

```
## Package Plan ##
```

```
environment location: /home/helena.despindula/.conda/envs/fsm-lite-conda
```

```
added / updated specs:
- conda-forge::sdsl-lite
```

```
The following packages will be downloaded:
```

package	build			
libgcc-15.1.0	h767d61c_4	805 KB	conda-forge	
libgcc-ng-15.1.0	h69a702a_4	29 KB	conda-forge	
libgomp-15.1.0	h767d61c_4	437 KB	conda-forge	
libstdcxx-15.1.0	h8f9b012_4	3.7 MB	conda-forge	
libstdcxx-ng-15.1.0	h4852527_4	29 KB	conda-forge	
Total:		5.0 MB		

```
The following NEW packages will be INSTALLED:
```

_libgcc_mutex	pkgs/main/linux-64::_libgcc_mutex-0.1-main
_openmp_mutex	pkgs/main/linux-64::_openmp_mutex-5.1-1_gnu
libgcc	conda-forge/linux-64::libgcc-15.1.0-h767d61c_4
libgcc-ng	conda-forge/linux-64::libgcc-ng-15.1.0-h69a702a_4
libgomp	conda-forge/linux-64::libgomp-15.1.0-h767d61c_4
libstdcxx	conda-forge/linux-64::libstdcxx-15.1.0-h8f9b012_4
libstdcxx-ng	conda-forge/linux-64::libstdcxx-ng-15.1.0-h4852527_4
sdsl-lite	conda-forge/linux-64::sdsl-lite-2.1.1-h00ab1b0_1002

```
Proceed ([y]/n)? y
```

```
Downloading and Extracting Packages:
```

```
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
```

```
(fsm-lite-conda) helena.despindula@BIOINF008:~$ conda install bioconda::fsm-lite
/usr/local/miniconda3/lib/python3.9/site-packages/conda/base/context.py:211: FutureWarning: Adding 'def
```

```
To remove this warning, please choose a default channel explicitly with conda's regular configuration sy
```

```
conda config --add channels defaults
```

```
For more information see https://docs.conda.io/projects/conda/en/stable/user-guide/configuration/use-con
```

```

    deprecated.topic(
Channels:
  - defaults
  - bioconda
Platform: linux-64
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##

environment location: /home/helena.despindula/.conda/envs/fsm-lite-conda

added / updated specs:
  - bioconda::fsm-lite

The following packages will be downloaded:

package | build
-----|-----
fsm-lite-1.0 | h9948957_6 251 KB bioconda
-----|-----
Total: 251 KB

The following NEW packages will be INSTALLED:

fsm-lite          bioconda/linux-64::fsm-lite-1.0-h9948957_6

Proceed ([y]/n)? y

Downloading and Extracting Packages:

Preparing transaction: done
Verifying transaction: done
Executing transaction: done
(fsm-lite-conda) helena.despindula@BIOINFO08:~$ fsm-li
fsm-li: command not found
(fsm-lite-conda) helena.despindula@BIOINFO08:~$ fsm-lite

(fsm-lite-conda) helena.despindula@BIOINFO08:~$ conda list --explicit
# This file may be used to create an environment using:
# $ conda create --name <env> --file <this file>
# platform: linux-64
# created-by: conda 25.5.1
@EXPLICIT
https://repo.anaconda.com/pkgs/main/linux-64/_libgcc_mutex-0.1-main.conda
https://conda.anaconda.org/conda-forge/linux-64/libgomp-15.1.0-h767d61c_4.conda

```

```
https://repo.anaconda.com/pkgs/main/linux-64/_openmp_mutex-5.1-1_gnu.conda
https://conda.anaconda.org/conda-forge/linux-64/libgcc-15.1.0-h767d61c_4.conda
https://conda.anaconda.org/conda-forge/linux-64/libgcc-ng-15.1.0-h69a702a_4.conda
https://conda.anaconda.org/conda-forge/linux-64/libstdcxx-15.1.0-h8f9b012_4.conda
https://conda.anaconda.org/conda-forge/linux-64/libstdcxx-ng-15.1.0-h4852527_4.conda
https://conda.anaconda.org/conda-forge/linux-64/sdsl-lite-2.1.1-h00ab1b0_1002.conda
https://conda.anaconda.org/bioconda/linux-64/fsm-lite-1.0-h9948957_6.tar.bz2
```

Teste de funcionamento:

```
(fsm-lite-conda) helena.despindula@BIOINFO08:~$ time fsm-lite -v --debug -m 2 -l LACTAS-HELISSON-01/He1
Reading input files...
Read 10 input files and 1154 sequences of total length 80325904 (includes rev.compl. sequences)
Constructing the data structures...
bwt end marker pos = 50209239
wt size = 80325905, n = 80325905
Construction complete, the main index requires 230.716 MiB plus 56.9204 MiB for labels.
All done.

real    6m56,624s
user    6m48,520s
sys     0m7,208s
```

Dessa vez o resultado não estava vazio.

Suspeito-se de um problema com as bibliotecas. Então tentamos reinstalar manualmente o `fsm-lite` no ambiente `conda fsm-lite-conda` com o Makefile original da versão 1.0 (com o caminho para o `sdsl-lite` do `conda`) e o comando `make clean && make depend && make`.

Makefile

```
SDSL_INSTALL_PREFIX=/home/helena.despindula/.conda/pkgs/sdsl-lite-2.1.1-h00ab1b0_1002

CPPFLAGS=-std=c++11 -I$(SDSL_INSTALL_PREFIX)/include -DNDEBUG -O3 -msse4.2
LIBS=-lsdsl -ldivsufsort -ldivsufsort64
OBJ = configuration.o input_reader.o fsm-lite.o

fsm-lite: $(OBJ)
    $(LINK.cpp) $^ -L$(SDSL_INSTALL_PREFIX)/lib $(LIBS) -o $@

test: fsm-lite
    ./fsm-lite -l test.list -t tmp -v --debug -m 1

clean:
    $(RM) fsm-lite *.o *~

depend:
    g++ -MM -std=c++11 -I$(SDSL_INSTALL_PREFIX)/include *.cpp > dependencies.mk

include dependencies.mk
```

Instalação:


```

(fsm-lite-conda) helena.despindula@BIOINFO08:~/LACTAS-HELISSON-01/Helena-stuff/fsm-lite/original$ make
rm -f fsm-lite *.o *~
(fsm-lite-conda) helena.despindula@BIOINFO08:~/LACTAS-HELISSON-01/Helena-stuff/fsm-lite/original$ make
g++ -MM -std=c++11 -I/home/helena.despindula/.conda/pkgs/sdsl-lite-2.1.1-h00ab1b0_1002/include *.cpp >
(fsm-lite-conda) helena.despindula@BIOINFO08:~/LACTAS-HELISSON-01/Helena-stuff/fsm-lite/original$ make
g++ -std=c++11 -I/home/helena.despindula/.conda/pkgs/sdsl-lite-2.1.1-h00ab1b0_1002/include -DNDEBUG -O3
g++ -std=c++11 -I/home/helena.despindula/.conda/pkgs/sdsl-lite-2.1.1-h00ab1b0_1002/include -DNDEBUG -O3
g++ -std=c++11 -I/home/helena.despindula/.conda/pkgs/sdsl-lite-2.1.1-h00ab1b0_1002/include -DNDEBUG -O3
g++ -std=c++11 -I/home/helena.despindula/.conda/pkgs/sdsl-lite-2.1.1-h00ab1b0_1002/include -DNDEBUG -O3
(fsm-lite-conda) helena.despindula@BIOINFO08:~/LACTAS-HELISSON-01/Helena-stuff/fsm-lite/original$ ./fsm
Reading input files...
Read 10 input files and 1154 sequences of total length 80325904 (includes rev.compl. sequences)
Constructing the data structures...
bwt end marker pos = 50209239
wt size = 80325905, n = 80325905
Construction complete, the main index requires 230.716 MiB plus 56.9204 MiB for labels.
All done.
(fsm-lite-conda) helena.despindula@BIOINFO08:~/LACTAS-HELISSON-01/Helena-stuff/fsm-lite/original$ ls -l
total 2,0G
-rw-rw-r-- 1 helena.despindula helena.despindula 4,8K abr 12 2016 configuration.cpp
-rw-rw-r-- 1 helena.despindula helena.despindula 557 abr 12 2016 configuration.h
-rw-rw-r-- 1 helena.despindula helena.despindula 21K ago 2 10:09 configuration.o
-rw-rw-r-- 1 helena.despindula helena.despindula 173 abr 12 2016 default.h
-rw-rw-r-- 1 helena.despindula helena.despindula 11K ago 2 10:09 dependencies.mk
-rwxrwxr-x 1 helena.despindula helena.despindula 801K ago 2 10:10 fsm-lite
-rw-rw-r-- 1 helena.despindula helena.despindula 6,9K jul 23 22:01 fsm-lite.cpp
-rw-rw-r-- 1 helena.despindula helena.despindula 479K ago 2 10:10 fsm-lite.o
-rw-rw-r-- 1 helena.despindula helena.despindula 5,6K abr 12 2016 input_reader.cpp
-rw-rw-r-- 1 helena.despindula helena.despindula 1,7K abr 12 2016 input_reader.h
-rw-rw-r-- 1 helena.despindula helena.despindula 38K ago 2 10:09 input_reader.o
-rw-rw-r-- 1 helena.despindula helena.despindula 35K abr 12 2016 LICENSE.md
-rw-rw-r-- 1 helena.despindula helena.despindula 533 ago 2 10:09 Makefile
-rw-rw-r-- 1 helena.despindula helena.despindula 484 abr 12 2016 Makefile-original
-rw-rw-r-- 1 helena.despindula helena.despindula 1,3K abr 12 2016 README.md
-rw-rw-r-- 1 helena.despindula helena.despindula 1,9G ago 2 10:19 resultados_conda.txt
-rw-rw-r-- 1 helena.despindula helena.despindula 250 ago 2 10:13 temp.meta
-rw-rw-r-- 1 helena.despindula helena.despindula 77M ago 2 10:13 temp.tmp

```

Foi realizado novo teste e o funcionamento dessa vez foi normal.

3.4 Criação de controle de versão no github (fork do original):

Nossa versão para modificações e testes foi depositada no github a partir de fork do original. <https://github.com/HelenaDEspindula/fsm-lite>

3.5 Criação de arquivo .sh para execução padronizada

Para fins de comparação de desempenho foram utilizadas as Parâmetros fixos: - ./fsm-lite uso do programa versão compilada local - --debug de modo debug - -v de modo verbose - -m 1 que indica que a sequência deve estar presente em pelo menos 1 genoma

E a combinações dos seguintes parâmetros variáveis (que tem maior chance de interferir no desempenho do programa): - Número de genomas: 5 10 25 50 100 250 - Parâmetro -S: 10 50 200 600 - Saída: .txt ou .txt.gz

Para avaliação utilizou-se o `/usr/bin/time` que a saber fornece:

- Command being timed:
- User time (seconds)
- System time (seconds)
- Percent of CPU this job got
- Elapsed (wall clock) time (h:mm:ss or m:ss)
- Average shared text size (kbytes)
- Average unshared data size (kbytes)
- Average stack size (kbytes)
- Average total size (kbytes)
- Maximum resident set size (kbytes)
- Average resident set size (kbytes)
- Major (requiring I/O) page faults
- Minor (reclaiming a frame) page faults
- Voluntary context switches
- Involuntary context switches
- Swaps
- File system inputs
- File system outputs
- Socket messages sent
- Socket messages received
- Signals delivered
- Page size (bytes)
- Exit status

Arquivo fsm-lite/original/progressivo.sh

```
#!/bin/bash

INPUT_FILE="/home/helena.despindula/LACTAS-HELISSEON-01/Helena-stuff/input_fsm-lite_OXA-23_OXA-24_temp.txt"
LISTA="/home/helena.despindula/LACTAS-HELISSEON-01/Helena-stuff/input_fsm-lite_OXA-23_OXA-24_todos.txt"
LOG_DIR="logs"
INTERVAL_MONITOR=30
GENOMAS=(5 10 25 50 100 250)
SMINUSCULO=6
SMAIUSCULO=(10 50 200 600)
MMINUSCULO=1
VERSION="1_0"
TMP_DIR="/home/helena.despindula/LACTAS-HELISSEON-01/Helena-stuff/fsm-lite-temp"
RES_DIR="/home/helena.despindula/LACTAS-HELISSEON-01/Helena-stuff/fsm-lite-results/${VERSION}"
TIMESTAMP=$(date +%Y-%m-%d_%H-%M-%S)

PROGRAMA="./fsm-lite"

# Criar pastas
mkdir -p "$LOG_DIR/monitor"
mkdir -p "$LOG_DIR/output"
mkdir -p "$TMP_DIR"
mkdir -p "$RES_DIR"
```

```

for N in "${GENOMAS[@]}"; do

    # Criar sublista
    head -n "$N" "$LISTA" > "$INPUT_FILE"

    for J in "${SMAIUSCULO[@]}"; do

        echo "====="

        TIMESTAMP=$(date +%Y-%m-%d_%H-%M-%S)
        MONITOR_LOG="$LOG_DIR/monitor/fsm_monitor_log_v${VERSION}_${N}genomas_${J}_max_TXT--${TIMESTAMP}.log"
        OUTPUT_LOG="$LOG_DIR/output/fsm_output_log_v${VERSION}_${N}genomas_${J}_max_TXT--${TIMESTAMP}.txt"
        TMP_FILES="$TMP_DIR/fsm_tmp_files_v${VERSION}_${N}genomas_${J}_max_TXT--${TIMESTAMP}"
        OUTPUT_RES="$RES_DIR/fsm_results_v${VERSION}_${N}genomas_${J}_max_TXT--${TIMESTAMP}.txt"

        echo "Rodando fsm-lite v${VERSION} saida TXT: para $N amostras com $J de maximo as ${TIMESTAMP}." >
        echo "Rodando fsm-lite v${VERSION} saida TXT: para $N amostras com $J de maximo as ${TIMESTAMP}." >
        echo -e "timestamp\tcpu\tmem\tvsz\ttrss" >> "$MONITOR_LOG"

        echo "Rodando fsm-lite v${VERSION} saida TXT: para $N amostras com $J de maximo as ${TIMESTAMP}." >

        echo "Rodando fsm-lite v${VERSION} saida TXT: para $N amostras com $J de maximo as ${TIMESTAMP}."

        # Executar fsm-lite em background
        ( /usr/bin/time -v "$PROGRAMA" -l "$INPUT_FILE" -s $SMINUSCULO -S $SMAIUSCULO -m $MMINUSCULO --debug
        FSM_PID=$!)

        echo "Monitorando PID: $FSM_PID"

        # Monitorar enquanto o processo estiver rodando
        while kill -0 "$FSM_PID" 2>/dev/null; do
            timestamp=$(date +%s)
            ps -p "$FSM_PID" -o %cpu,%mem,vsz,rss --no-headers | \
                awk -v t="$timestamp" '{print t"\t"$1"\t"$2"\t"$3"\t"$4}' >> "$MONITOR_LOG"
            pidstat -h -r -u -p $FSM_PID 1 1 | awk -v t="$timestamp" 'NR==4 {print t"\t"$8"\t"$9"\t"$10"\t"$11}' >> "$MONITOR_LOG"
            sleep "$INTERVAL_MONITOR"
        done

        echo "====="

        TIMESTAMP=$(date +%Y-%m-%d_%H-%M-%S)
        MONITOR_LOG="$LOG_DIR/monitor/fsm_monitor_log_v${VERSION}_${N}genomas_${J}_max_GZ--${TIMESTAMP}.log"
        OUTPUT_LOG="$LOG_DIR/output/fsm_output_log_v${VERSION}_${N}genomas_${J}_max_GZ--${TIMESTAMP}.txt"
        TMP_FILES="$TMP_DIR/fsm_tmp_files_v${VERSION}_${N}genomas_${J}_max_GZ--${TIMESTAMP}"
        OUTPUT_RES="$RES_DIR/fsm_results_v${VERSION}_${N}genomas_${J}_max_GZ--${TIMESTAMP}.txt.gz"

        echo "Rodando fsm-lite v${VERSION} saida GZ: para $N amostras com $J de maximo as ${TIMESTAMP}." >
        echo "Rodando fsm-lite v${VERSION} saida GZ: para $N amostras com $J de maximo as ${TIMESTAMP}." >
        echo -e "timestamp\tcpu\tmem\tvsz\ttrss" >> "$MONITOR_LOG"
    done
done

```

```

echo "Rodando fsm-lite v${VERSION} saida GZ: para $N amostras com $J de maximo as ${TIMESTAMP}." >

echo "Rodando fsm-lite v${VERSION} saida GZ: para $N amostras com $J de maximo as ${TIMESTAMP}."

# Executar fsm-lite em background
( ( /usr/bin/time -v "$PROGRAMA" -l "$INPUT_FILE" -s $SMINUSCULO -S $SMAIUSCULO -m $MMINUSCULO --del
FSM_PID=$!

echo "Monitorando PID: $FSM_PID"

# Monitorar enquanto o processo estiver rodando
while kill -0 "$FSM_PID" 2>/dev/null; do
    timestamp=$(date +%s)
    ps -p "$FSM_PID" -o %cpu,%mem,vsz,rss --no-headers | \
        awk -v t="$timestamp" '{print t"\t"$1"\t"$2"\t"$3"\t"$4}' >> "$MONITOR_LOG"
    pidstat -h -r -u -p $FSM_PID 1 1 | awk -v t="$timestamp" 'NR==4 {print t"\t"$8"\t"$9"\t"$10"\t"$11}'
    sleep "$INTERVAL_MONITOR"
done

done

wait "$FSM_PID"
echo "Finalizado testes com $N amostras as ${TIMESTAMP}."
done

echo "====="
echo "Todos os testes foram concluídos."

```

Exemplo de log de output:

```

Reading input files...
Read 5 input files and 326 sequences of total length 39963630 (includes rev.compl. sequences)
Constructing the data structures...
bwt end marker pos = 24983797
wt size = 39963631, n = 39963631
Construction complete, the main index requires 96.3672 MiB plus 21.2455 MiB for labels.
All done.
Command being timed: "./fsm-lite -l /home/helena.despindula/LACTAS-HELISSON-01/Helena-stuff/input_f
User time (seconds): 61.55
System time (seconds): 1.34
Percent of CPU this job got: 99%
Elapsed (wall clock) time (h:mm:ss or m:ss): 1:03.37
Average shared text size (kbytes): 0
Average unshared data size (kbytes): 0
Average stack size (kbytes): 0
Average total size (kbytes): 0
Maximum resident set size (kbytes): 364996
Average resident set size (kbytes): 0
Major (requiring I/O) page faults: 0
Minor (reclaiming a frame) page faults: 224190
Voluntary context switches: 502
Involuntary context switches: 6113
Swaps: 0

```

```
File system inputs: 0
File system outputs: 785648
Socket messages sent: 0
Socket messages received: 0
Signals delivered: 0
Page size (bytes): 4096
Exit status: 0
```

Exemplo de Head de resultado

3.6 Avaliação dos logs de output

Tabela:

```
# Caminho dos logs
log_dir <- "original/logs/output/"
log_files <- list.files(log_dir, pattern = "fsm_output_log_.*\\.txt$", full.names = TRUE)

print("Foram encontrados os arquivos")
```

```
## [1] "Foram encontrados os arquivos"
```

```
print(log_files)
```

```
## [1] "original/logs/output/fsm_output_log_v1_0_100genomas_10_max_GZ--2025-08-04_05-25-09.txt"
## [2] "original/logs/output/fsm_output_log_v1_0_100genomas_10_max_TXT--2025-08-04_05-00-14.txt"
## [3] "original/logs/output/fsm_output_log_v1_0_100genomas_200_max_GZ--2025-08-04_07-05-17.txt"
## [4] "original/logs/output/fsm_output_log_v1_0_100genomas_200_max_TXT--2025-08-04_06-39-51.txt"
## [5] "original/logs/output/fsm_output_log_v1_0_100genomas_50_max_GZ--2025-08-04_06-14-57.txt"
## [6] "original/logs/output/fsm_output_log_v1_0_100genomas_50_max_TXT--2025-08-04_05-50-03.txt"
## [7] "original/logs/output/fsm_output_log_v1_0_100genomas_600_max_GZ--2025-08-04_07-55-37.txt"
## [8] "original/logs/output/fsm_output_log_v1_0_100genomas_600_max_TXT--2025-08-04_07-30-11.txt"
## [9] "original/logs/output/fsm_output_log_v1_0_10genomas_10_max_GZ--2025-08-04_01-37-52.txt"
## [10] "original/logs/output/fsm_output_log_v1_0_10genomas_10_max_TXT--2025-08-04_01-32-41.txt"
## [11] "original/logs/output/fsm_output_log_v1_0_10genomas_200_max_GZ--2025-08-04_01-57-05.txt"
## [12] "original/logs/output/fsm_output_log_v1_0_10genomas_200_max_TXT--2025-08-04_01-52-24.txt"
## [13] "original/logs/output/fsm_output_log_v1_0_10genomas_50_max_GZ--2025-08-04_01-47-44.txt"
## [14] "original/logs/output/fsm_output_log_v1_0_10genomas_50_max_TXT--2025-08-04_01-43-04.txt"
## [15] "original/logs/output/fsm_output_log_v1_0_10genomas_600_max_GZ--2025-08-04_02-07-27.txt"
## [16] "original/logs/output/fsm_output_log_v1_0_10genomas_600_max_TXT--2025-08-04_02-02-47.txt"
## [17] "original/logs/output/fsm_output_log_v1_0_25genomas_10_max_GZ--2025-08-04_02-18-21.txt"
## [18] "original/logs/output/fsm_output_log_v1_0_25genomas_10_max_TXT--2025-08-04_02-12-07.txt"
## [19] "original/logs/output/fsm_output_log_v1_0_25genomas_200_max_GZ--2025-08-04_02-43-15.txt"
## [20] "original/logs/output/fsm_output_log_v1_0_25genomas_200_max_TXT--2025-08-04_02-37-02.txt"
## [21] "original/logs/output/fsm_output_log_v1_0_25genomas_50_max_GZ--2025-08-04_02-30-48.txt"
## [22] "original/logs/output/fsm_output_log_v1_0_25genomas_50_max_TXT--2025-08-04_02-24-35.txt"
## [23] "original/logs/output/fsm_output_log_v1_0_25genomas_600_max_GZ--2025-08-04_02-55-43.txt"
## [24] "original/logs/output/fsm_output_log_v1_0_25genomas_600_max_TXT--2025-08-04_02-49-29.txt"
## [25] "original/logs/output/fsm_output_log_v1_0_50genomas_10_max_GZ--2025-08-04_03-16-28.txt"
## [26] "original/logs/output/fsm_output_log_v1_0_50genomas_10_max_TXT--2025-08-04_03-01-56.txt"
```

```
## [27] "original/logs/output/fsm_output_log_v1_0_50genomas_200_max_GZ--2025-08-04_04-15-37.txt"
## [28] "original/logs/output/fsm_output_log_v1_0_50genomas_200_max_TXT--2025-08-04_04-01-05.txt"
## [29] "original/logs/output/fsm_output_log_v1_0_50genomas_50_max_GZ--2025-08-04_03-46-02.txt"
## [30] "original/logs/output/fsm_output_log_v1_0_50genomas_50_max_TXT--2025-08-04_03-31-31.txt"
## [31] "original/logs/output/fsm_output_log_v1_0_50genomas_600_max_GZ--2025-08-04_04-45-43.txt"
## [32] "original/logs/output/fsm_output_log_v1_0_50genomas_600_max_TXT--2025-08-04_04-30-40.txt"
## [33] "original/logs/output/fsm_output_log_v1_0_5genomas_10_max_GZ--2025-08-04_01-23-52.txt"
## [34] "original/logs/output/fsm_output_log_v1_0_5genomas_10_max_TXT--2025-08-04_01-22-18.txt"
## [35] "original/logs/output/fsm_output_log_v1_0_5genomas_200_max_GZ--2025-08-04_01-29-03.txt"
## [36] "original/logs/output/fsm_output_log_v1_0_5genomas_200_max_TXT--2025-08-04_01-27-30.txt"
## [37] "original/logs/output/fsm_output_log_v1_0_5genomas_50_max_GZ--2025-08-04_01-26-27.txt"
## [38] "original/logs/output/fsm_output_log_v1_0_5genomas_50_max_TXT--2025-08-04_01-25-25.txt"
## [39] "original/logs/output/fsm_output_log_v1_0_5genomas_600_max_GZ--2025-08-04_01-31-08.txt"
## [40] "original/logs/output/fsm_output_log_v1_0_5genomas_600_max_TXT--2025-08-04_01-30-05.txt"
```

```
tempo_para_segundos <- function(t) {
  if (grepl(":", t)) {
    partes <- as.numeric(unlist(strsplit(t, ":")))
    if (length(partes) == 3) return(partes[1]*3600 + partes[2]*60 + partes[3])
    if (length(partes) == 2) return(partes[1]*60 + partes[2])
  }
  return(as.numeric(t))
}
```

```
extrair_dados_log <- function(arquivo) {
  linhas <- readLines(arquivo)
  nome <- basename(arquivo)

  # Extrair metadados do nome do arquivo
  versao <- str_match(nome, "fsm_output_log_(v[0-9_]+)")[,2]
  genomas <- as.numeric(str_match(nome, "_([0-9_]+)genomas")[,2])
  max_param <- as.numeric(str_match(nome, "_([0-9_]+)_max")[,2])
  tipo_output <- str_match(nome, "_([TXT|GZ])--")[,2]
  timestamp <- str_match(nome, "--([0-9]{4}-[0-9]{2}-[0-9]{2}_[0-9-]{8})")[,2]

  get_valor <- function(chave) {
    linha <- grep(chave, linhas, value = TRUE)
    if (length(linha) > 0) {
      return(str_trim(gsub(".*:\\s*", "", linha)))
    } else {
      return(NA)
    }
  }

  # Extrai tempo decorrido como string tipo "4:31.51"
  get_elapsed <- function(linhas) {
    linha <- grep("Elapsed \\(wall clock\\) time", linhas, value = TRUE)
    if (length(linha) > 0) {
      tempo <- str_match(linha, ":(\\s*(\\d+:\\d+(\\.\\d+)?))")[,2]
      return(tempo)
    } else {
      return(NA)
    }
  }
}
```

```

# Extração dos dados do conteúdo
elapsed <- tempo_para_segundos(get_elapsed(linhas))
user <- as.numeric(get_valor("User time"))
system <- as.numeric(get_valor("System time"))
cpu_pct <- as.numeric(gsub("%", "", get_valor("Percent of CPU")))
max_mem_kb <- as.numeric(get_valor("Maximum resident set size"))
max_mem_mb <- round(max_mem_kb / 1024, 2)

#print(elapsed)

data.frame(
  arquivo = nome,
  versao = versao,
  genomas = genomas,
  max_param = max_param,
  tipo_output = tipo_output,
  timestamp = timestamp,
  elapsed_sec = elapsed,
  user_time = user,
  system_time = system,
  cpu_percent = cpu_pct,
  max_rss_mb = max_mem_mb
)
}

# Aplicar a função
tabela_logs <- bind_rows(lapply(log_files, extrair_dados_log))

# Visualizar
print(tabela_logs)

```

```

##                                     arquivo  versao
## 1 fsm_output_log_v1_0_100genomas_10_max_GZ--2025-08-04_05-25-09.txt v1_0_100
## 2 fsm_output_log_v1_0_100genomas_10_max_TXT--2025-08-04_05-00-14.txt v1_0_100
## 3 fsm_output_log_v1_0_100genomas_200_max_GZ--2025-08-04_07-05-17.txt v1_0_100
## 4 fsm_output_log_v1_0_100genomas_200_max_TXT--2025-08-04_06-39-51.txt v1_0_100
## 5 fsm_output_log_v1_0_100genomas_50_max_GZ--2025-08-04_06-14-57.txt v1_0_100
## 6 fsm_output_log_v1_0_100genomas_50_max_TXT--2025-08-04_05-50-03.txt v1_0_100
## 7 fsm_output_log_v1_0_100genomas_600_max_GZ--2025-08-04_07-55-37.txt v1_0_100
## 8 fsm_output_log_v1_0_100genomas_600_max_TXT--2025-08-04_07-30-11.txt v1_0_100
## 9 fsm_output_log_v1_0_10genomas_10_max_GZ--2025-08-04_01-37-52.txt v1_0_10
## 10 fsm_output_log_v1_0_10genomas_10_max_TXT--2025-08-04_01-32-41.txt v1_0_10
## 11 fsm_output_log_v1_0_10genomas_200_max_GZ--2025-08-04_01-57-05.txt v1_0_10
## 12 fsm_output_log_v1_0_10genomas_200_max_TXT--2025-08-04_01-52-24.txt v1_0_10
## 13 fsm_output_log_v1_0_10genomas_50_max_GZ--2025-08-04_01-47-44.txt v1_0_10
## 14 fsm_output_log_v1_0_10genomas_50_max_TXT--2025-08-04_01-43-04.txt v1_0_10
## 15 fsm_output_log_v1_0_10genomas_600_max_GZ--2025-08-04_02-07-27.txt v1_0_10
## 16 fsm_output_log_v1_0_10genomas_600_max_TXT--2025-08-04_02-02-47.txt v1_0_10
## 17 fsm_output_log_v1_0_25genomas_10_max_GZ--2025-08-04_02-18-21.txt v1_0_25
## 18 fsm_output_log_v1_0_25genomas_10_max_TXT--2025-08-04_02-12-07.txt v1_0_25
## 19 fsm_output_log_v1_0_25genomas_200_max_GZ--2025-08-04_02-43-15.txt v1_0_25
## 20 fsm_output_log_v1_0_25genomas_200_max_TXT--2025-08-04_02-37-02.txt v1_0_25
## 21 fsm_output_log_v1_0_25genomas_50_max_GZ--2025-08-04_02-30-48.txt v1_0_25

```



```

## 22 fsm_output_log_v1_0_25genomas_50_max_TXT--2025-08-04_02-24-35.txt v1_0_25
## 23 fsm_output_log_v1_0_25genomas_600_max_GZ--2025-08-04_02-55-43.txt v1_0_25
## 24 fsm_output_log_v1_0_25genomas_600_max_TXT--2025-08-04_02-49-29.txt v1_0_25
## 25 fsm_output_log_v1_0_50genomas_10_max_GZ--2025-08-04_03-16-28.txt v1_0_50
## 26 fsm_output_log_v1_0_50genomas_10_max_TXT--2025-08-04_03-01-56.txt v1_0_50
## 27 fsm_output_log_v1_0_50genomas_200_max_GZ--2025-08-04_04-15-37.txt v1_0_50
## 28 fsm_output_log_v1_0_50genomas_200_max_TXT--2025-08-04_04-01-05.txt v1_0_50
## 29 fsm_output_log_v1_0_50genomas_50_max_GZ--2025-08-04_03-46-02.txt v1_0_50
## 30 fsm_output_log_v1_0_50genomas_50_max_TXT--2025-08-04_03-31-31.txt v1_0_50
## 31 fsm_output_log_v1_0_50genomas_600_max_GZ--2025-08-04_04-45-43.txt v1_0_50
## 32 fsm_output_log_v1_0_50genomas_600_max_TXT--2025-08-04_04-30-40.txt v1_0_50
## 33 fsm_output_log_v1_0_5genomas_10_max_GZ--2025-08-04_01-23-52.txt v1_0_5
## 34 fsm_output_log_v1_0_5genomas_10_max_TXT--2025-08-04_01-22-18.txt v1_0_5
## 35 fsm_output_log_v1_0_5genomas_200_max_GZ--2025-08-04_01-29-03.txt v1_0_5
## 36 fsm_output_log_v1_0_5genomas_200_max_TXT--2025-08-04_01-27-30.txt v1_0_5
## 37 fsm_output_log_v1_0_5genomas_50_max_GZ--2025-08-04_01-26-27.txt v1_0_5
## 38 fsm_output_log_v1_0_5genomas_50_max_TXT--2025-08-04_01-25-25.txt v1_0_5
## 39 fsm_output_log_v1_0_5genomas_600_max_GZ--2025-08-04_01-31-08.txt v1_0_5
## 40 fsm_output_log_v1_0_5genomas_600_max_TXT--2025-08-04_01-30-05.txt v1_0_5
##      genomas max_param tipo_output      timestamp elapsed_sec user_time
## 1      100          10          GZ 2025-08-04_05-25-09      1474.04    1430.14
## 2      100          10          TXT 2025-08-04_05-00-14      1480.63    1435.44
## 3      100         200          GZ 2025-08-04_07-05-17      1492.00    1437.98
## 4      100         200          TXT 2025-08-04_06-39-51      1495.59    1435.16
## 5      100          50          GZ 2025-08-04_06-14-57      1486.90    1442.17
## 6      100          50          TXT 2025-08-04_05-50-03      1484.34    1440.23
## 7      100         600          GZ 2025-08-04_07-55-37      1504.46    1455.12
## 8      100         600          TXT 2025-08-04_07-30-11      1501.68    1440.61
## 9       10          10          GZ 2025-08-04_01-37-52       302.45     297.49
## 10      10          10          TXT 2025-08-04_01-32-41       300.91     293.76
## 11      10         200          GZ 2025-08-04_01-57-05       314.47     310.03
## 12      10         200          TXT 2025-08-04_01-52-24       270.40     263.46
## 13      10          50          GZ 2025-08-04_01-47-44       271.51     267.00
## 14      10          50          TXT 2025-08-04_01-43-04       268.27     260.93
## 15      10         600          GZ 2025-08-04_02-07-27       269.22     264.82
## 16      10         600          TXT 2025-08-04_02-02-47       280.01     273.05
## 17      25          10          GZ 2025-08-04_02-18-21       348.22     338.76
## 18      25          10          TXT 2025-08-04_02-12-07       350.18     340.13
## 19      25         200          GZ 2025-08-04_02-43-15       352.08     341.76
## 20      25         200          TXT 2025-08-04_02-37-02       359.97     349.29
## 21      25          50          GZ 2025-08-04_02-30-48       366.31     356.18
## 22      25          50          TXT 2025-08-04_02-24-35       347.99     338.93
## 23      25         600          GZ 2025-08-04_02-55-43       349.18     338.81
## 24      25         600          TXT 2025-08-04_02-49-29       349.07     339.41
## 25      50          10          GZ 2025-08-04_03-16-28       892.64     871.57
## 26      50          10          TXT 2025-08-04_03-01-56       864.85     843.08
## 27      50         200          GZ 2025-08-04_04-15-37       896.31     877.29
## 28      50         200          TXT 2025-08-04_04-01-05       861.22     839.73
## 29      50          50          GZ 2025-08-04_03-46-02       882.10     858.15
## 30      50          50          TXT 2025-08-04_03-31-31       864.47     844.00
## 31      50         600          GZ 2025-08-04_04-45-43       861.85     841.20
## 32      50         600          TXT 2025-08-04_04-30-40       899.71     876.52
## 33       5          10          GZ 2025-08-04_01-23-52        63.37      61.55
## 34       5          10          TXT 2025-08-04_01-22-18        63.83      61.85

```


## 35	5	200	GZ 2025-08-04_01-29-03	61.66	59.92
## 36	5	200	TXT 2025-08-04_01-27-30	63.97	62.18
## 37	5	50	GZ 2025-08-04_01-26-27	62.08	59.83
## 38	5	50	TXT 2025-08-04_01-25-25	61.17	59.17
## 39	5	600	GZ 2025-08-04_01-31-08	63.94	62.06
## 40	5	600	TXT 2025-08-04_01-30-05	60.96	59.08
##	system_time	cpu_percent	max_rss_mb		
## 1	36.96	99	7734.21		
## 2	36.94	99	7734.09		
## 3	38.02	98	7734.08		
## 4	42.12	98	7734.14		
## 5	37.70	99	7734.15		
## 6	35.44	99	7734.15		
## 7	40.51	99	7734.16		
## 8	38.52	98	7734.44		
## 9	3.58	99	748.07		
## 10	4.50	99	748.01		
## 11	3.40	99	748.07		
## 12	4.58	99	748.04		
## 13	3.60	99	748.07		
## 14	4.61	98	748.19		
## 15	3.44	99	747.95		
## 16	4.48	99	748.04		
## 17	7.49	99	1794.68		
## 18	8.04	99	1794.50		
## 19	8.54	99	1794.45		
## 20	8.05	99	1794.52		
## 21	8.03	99	1794.47		
## 22	7.30	99	1794.58		
## 23	8.58	99	1794.46		
## 24	7.71	99	1794.60		
## 25	17.87	99	3835.08		
## 26	18.22	99	3835.20		
## 27	15.96	99	3835.14		
## 28	18.38	99	3835.17		
## 29	20.75	99	3835.30		
## 30	17.24	99	3835.14		
## 31	17.59	99	3835.29		
## 32	19.52	99	3835.32		
## 33	1.34	99	356.44		
## 34	1.52	99	356.31		
## 35	1.29	99	356.39		
## 36	1.34	99	356.53		
## 37	1.75	99	356.33		
## 38	1.51	99	356.38		
## 39	1.41	99	356.37		
## 40	1.43	99	356.52		

```
#View(tabela_logs)
```

```
# Exportar se quiser
```

```
write.csv(tabela_logs, "original/resumo_execucoes_v1_0.csv", row.names = FALSE)
```

Graficos:

```

# 1. Ler a tabela
tabela_logs <- read.csv("original/resumo_execucoes_v1_0.csv")

# 2. Garantir tipos corretos
tabela_logs$tipo_output <- factor(tabela_logs$tipo_output, levels = c("TXT", "GZ"))
tabela_logs$genomas <- as.numeric(tabela_logs$genomas)
tabela_logs$max_param <- as.numeric(tabela_logs$max_param)

# 3. Criar pasta para salvar
dir.create("graficos", showWarnings = FALSE)

# Função auxiliar para salvar
salvar_plot <- function(p, nome, width = 9, height = 6) {
  ggsave(filename = paste0("graficos/", nome, ".png"), plot = p, width = width, height = height, dpi = 300)
}

# 4. Geração dos gráficos

## 1. Tempo vs max_param (fixando genomas)
p1 <- ggplot(tabela_logs, aes(x = max_param, y = elapsed_sec, color = tipo_output)) +
  geom_line(aes(group = interaction(tipo_output, genomas)), linewidth = 1) +
  geom_point(size = 2) +
  facet_wrap(~genomas, scales = "free_y", labeller = label_both) +
  labs(title = "Tempo vs max_param (fixando número de genomas)",
       x = "Parâmetro máximo", y = "Tempo decorrido (s)", color = "Saída")
p1
salvar_plot(p1, "tempo_vs_maxparam_por_genoma")

## 2. Memória vs max_param (fixando genomas)
p2 <- ggplot(tabela_logs, aes(x = max_param, y = max_rss_mb, color = tipo_output)) +
  geom_line(aes(group = interaction(tipo_output, genomas)), linewidth = 1) +
  geom_point(size = 2) +
  facet_wrap(~genomas, scales = "free_y", labeller = label_both) +
  labs(title = "Uso de memória vs max_param (fixando número de genomas)",
       x = "Parâmetro máximo", y = "Pico de RAM (MB)", color = "Saída")
p2
salvar_plot(p2, "memoria_vs_maxparam_por_genoma")

## 3. Tempo vs genomas (fixando max_param)
p3 <- ggplot(tabela_logs, aes(x = genomas, y = elapsed_sec, color = tipo_output)) +
  geom_line(aes(group = interaction(tipo_output, max_param)), linewidth = 1) +
  geom_point(size = 2) +
  facet_wrap(~max_param, scales = "free_y", labeller = label_both) +
  labs(title = "Tempo vs número de genomas (fixando max_param)",
       x = "Número de genomas", y = "Tempo decorrido (s)", color = "Saída")
p3
salvar_plot(p3, "tempo_vs_genomas_por_maxparam")

## 4. Memória vs genomas (fixando max_param)
p4 <- ggplot(tabela_logs, aes(x = genomas, y = max_rss_mb, color = tipo_output)) +
  geom_line(aes(group = interaction(tipo_output, max_param)), linewidth = 1) +
  geom_point(size = 2) +
  facet_wrap(~max_param, scales = "free_y", labeller = label_both) +

```

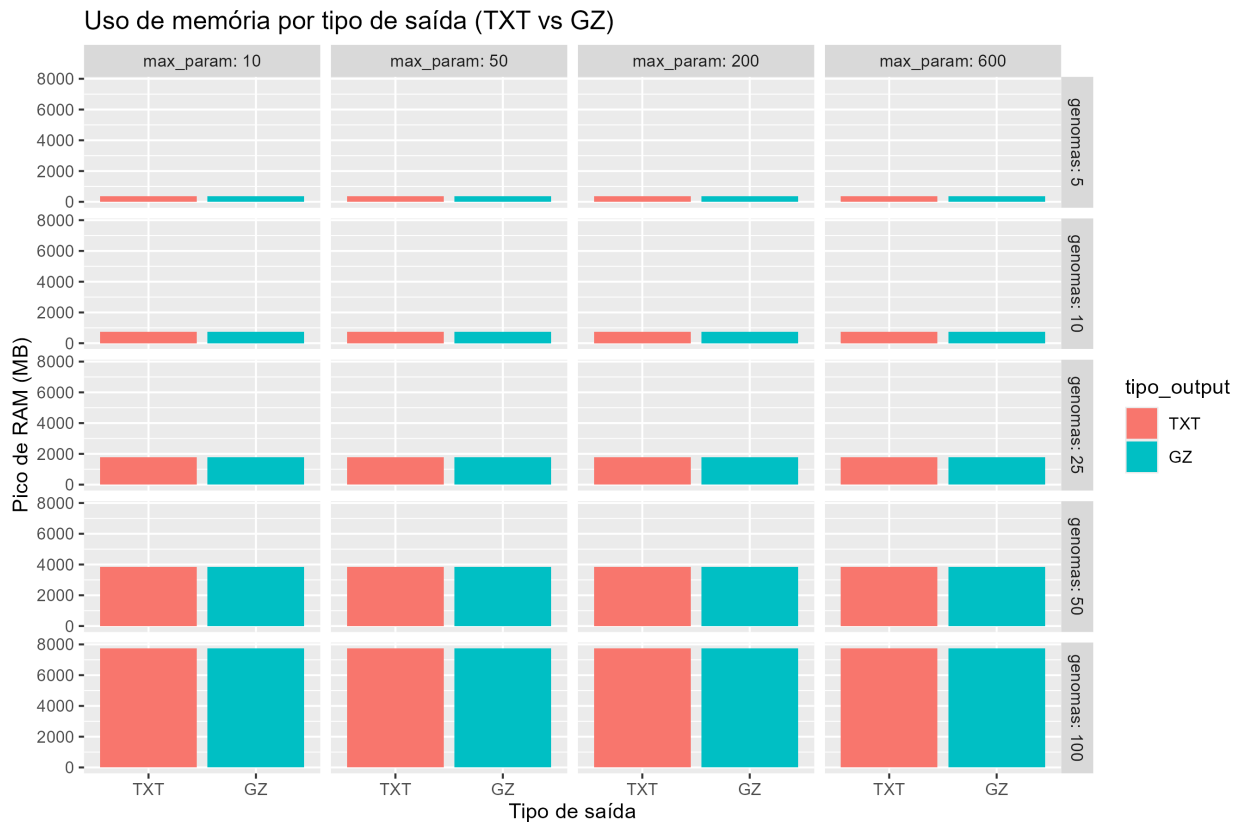
```

labs(title = "Uso de memória vs número de genomas (fixando max_param)",
     x = "Número de genomas", y = "Pico de RAM (MB)", color = "Saída")
p4
salvar_plot(p4, "memoria_vs_genomas_por_maxparam")

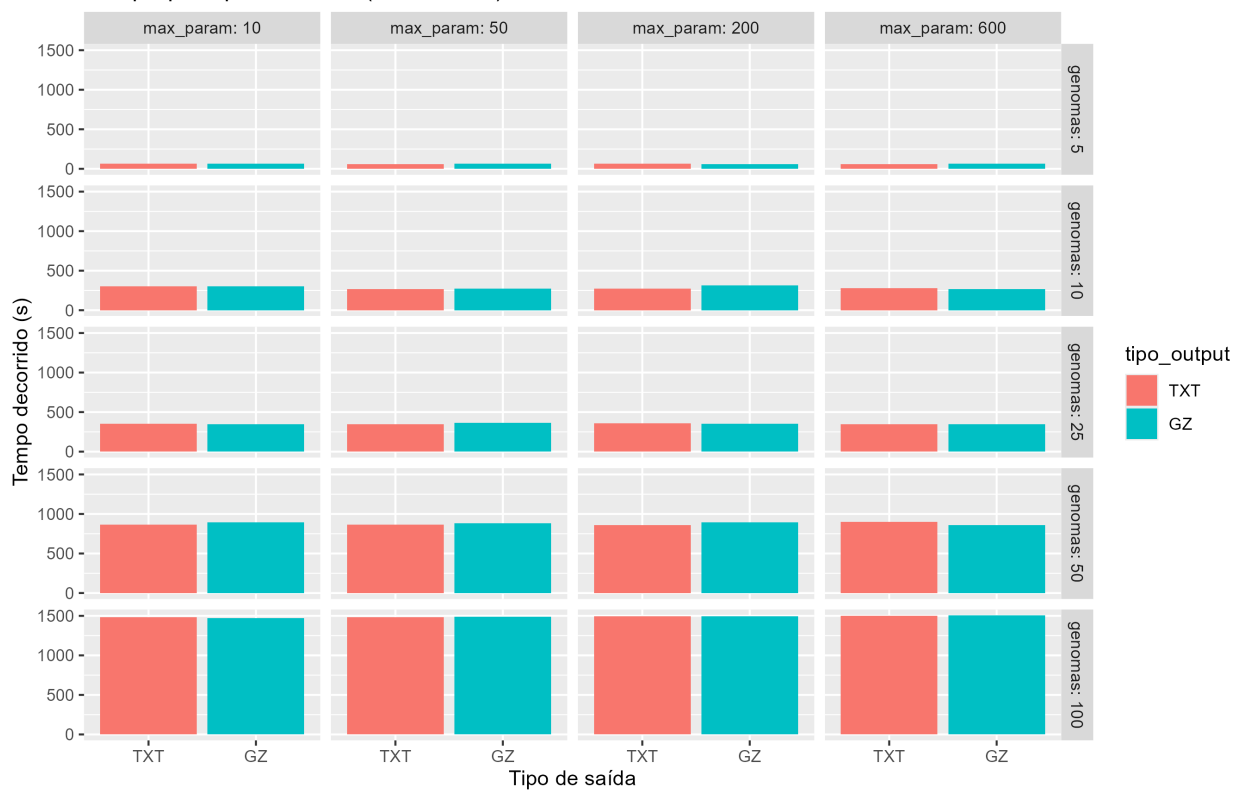
## 5. Comparação TXT vs GZ (barras por faceta)
p5 <- ggplot(tabela_logs, aes(x = tipo_output, y = elapsed_sec, fill = tipo_output)) +
  geom_col(position = position_dodge()) +
  facet_grid(genomas ~ max_param, labeller = label_both) +
  labs(title = "Tempo por tipo de saída (TXT vs GZ)",
       x = "Tipo de saída", y = "Tempo decorrido (s)")
p5
salvar_plot(p5, "barras_tempo_tipo_saida")

p6 <- ggplot(tabela_logs, aes(x = tipo_output, y = max_rss_mb, fill = tipo_output)) +
  geom_col(position = position_dodge()) +
  facet_grid(genomas ~ max_param, labeller = label_both) +
  labs(title = "Uso de memória por tipo de saída (TXT vs GZ)",
       x = "Tipo de saída", y = "Pico de RAM (MB)")
p6
salvar_plot(p6, "barras_memoria_tipo_saida")

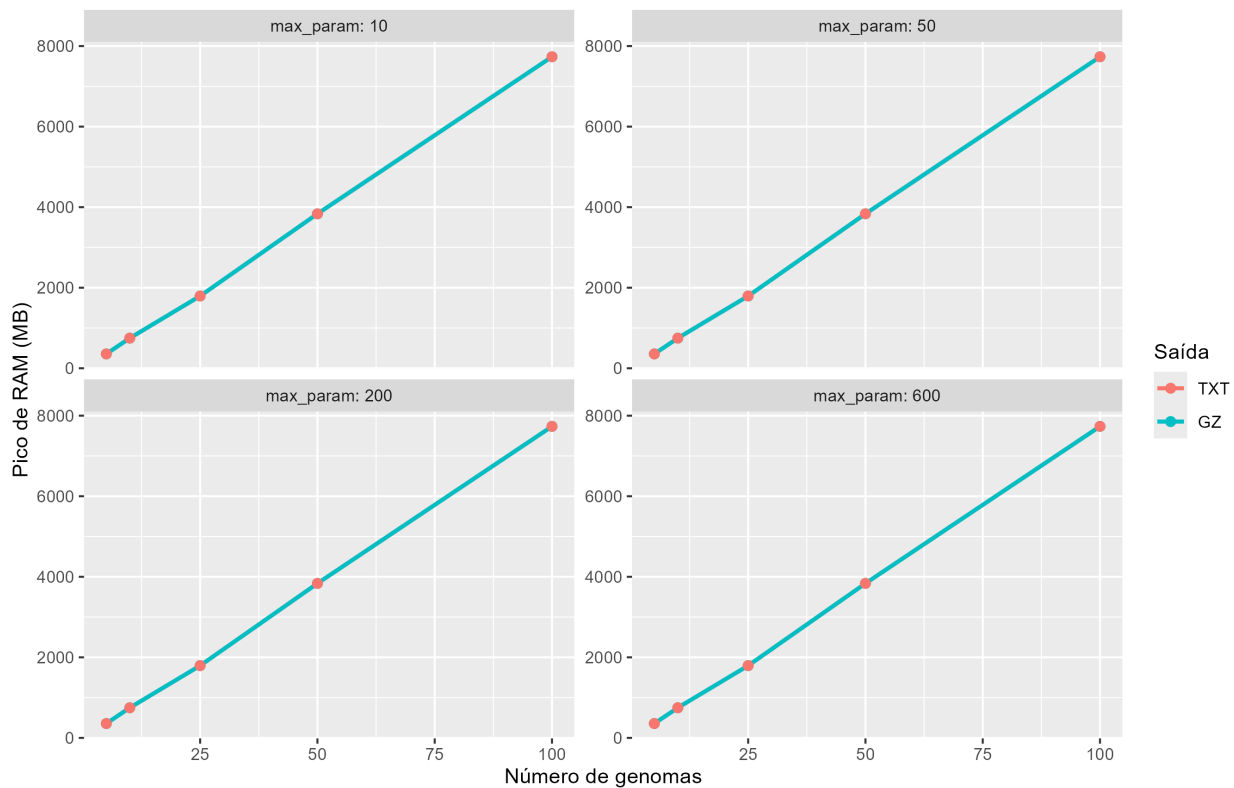
```



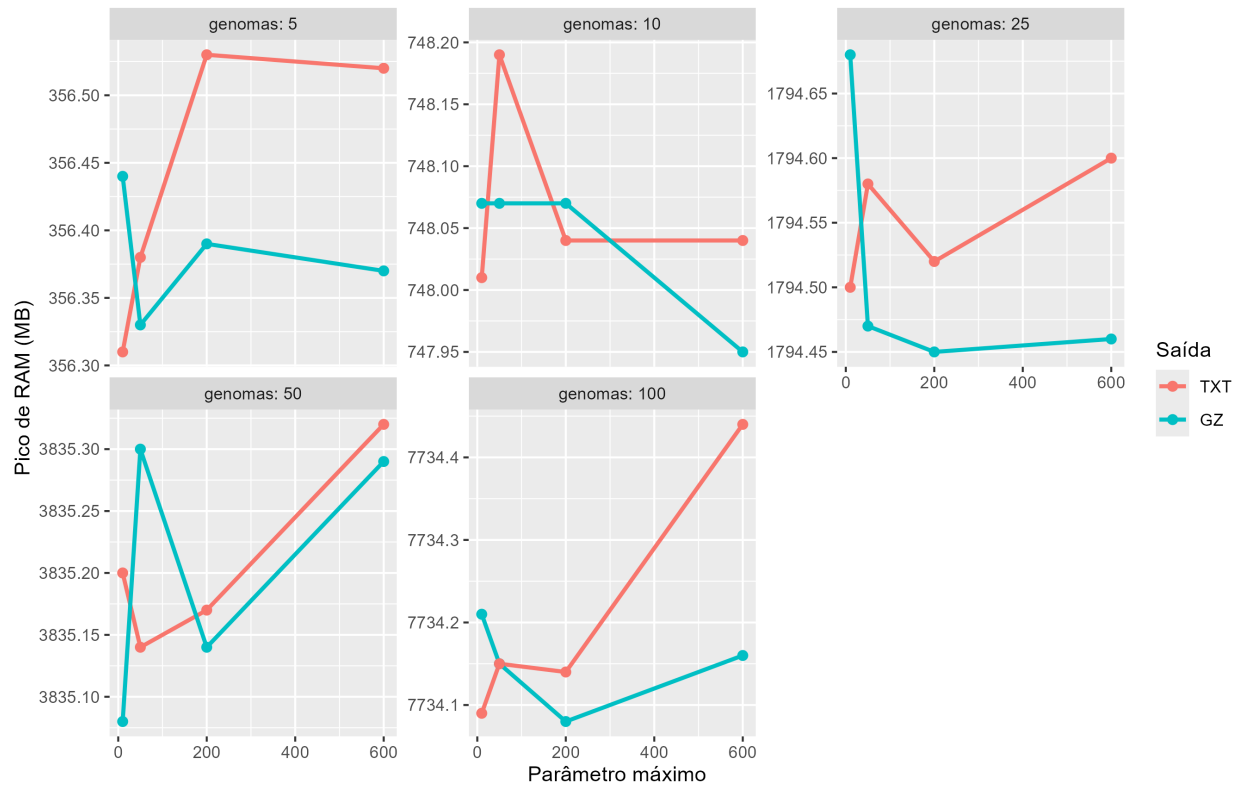
Tempo por tipo de saída (TXT vs GZ)



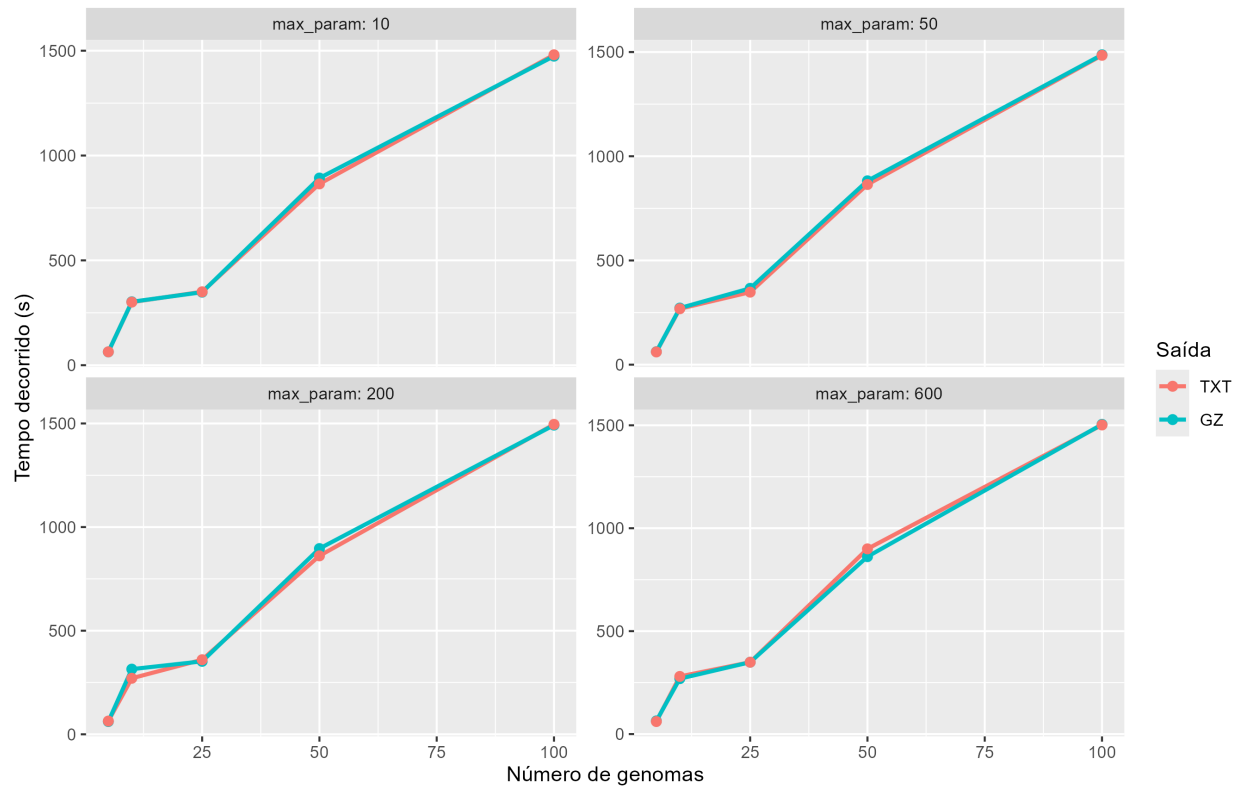
Uso de memória vs número de genomas (fixando max_param)



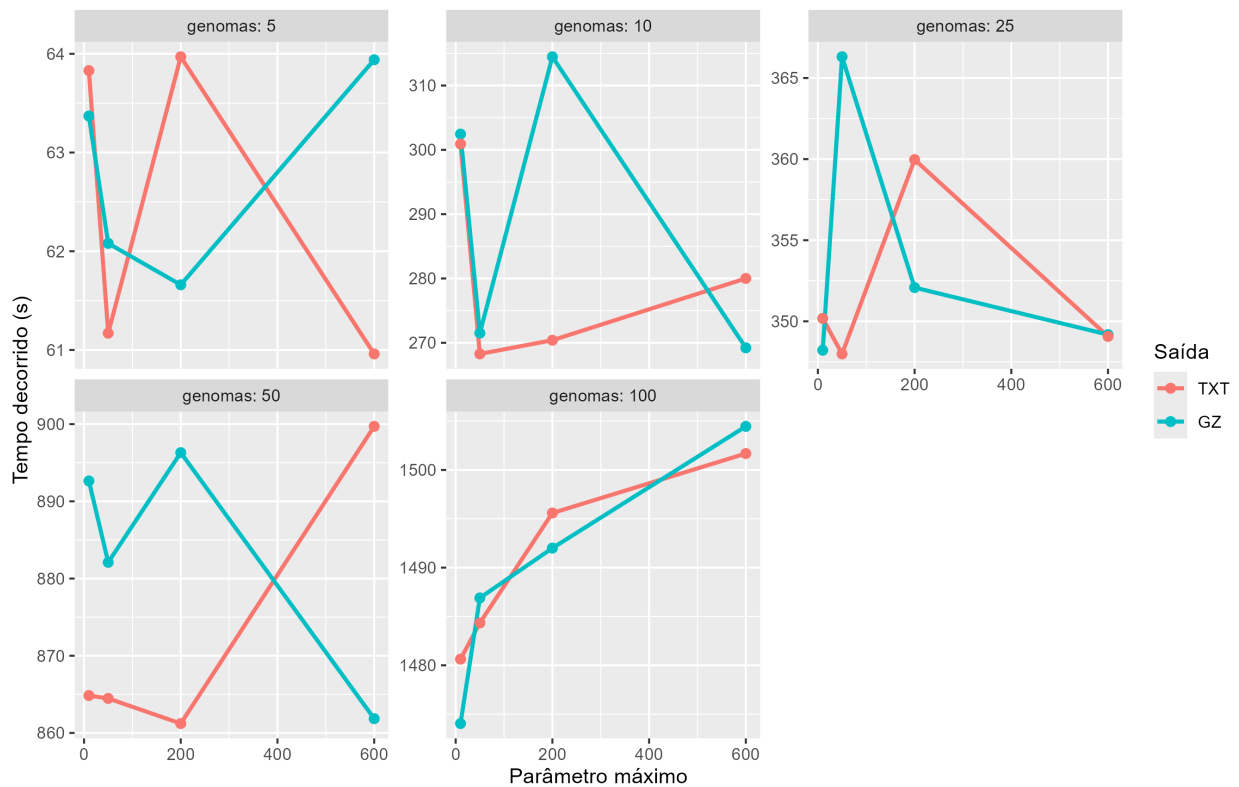
Uso de memória vs max_param (fixando número de genomas)



Tempo vs número de genomas (fixando max_param)



Tempo vs max_param (fixando número de genomas)



```
# Vamos pegar os dados de uma condição específica de max_param (ex: 200)
dados_modelo <- tabela_logs %>% filter(max_param == 200 & tipo_output == "TXT")

# Ajustar modelos
modelo_linear <- lm(elapsed_sec ~ genomas, data = dados_modelo)
modelo_quad <- lm(elapsed_sec ~ poly(genomas, 2), data = dados_modelo)

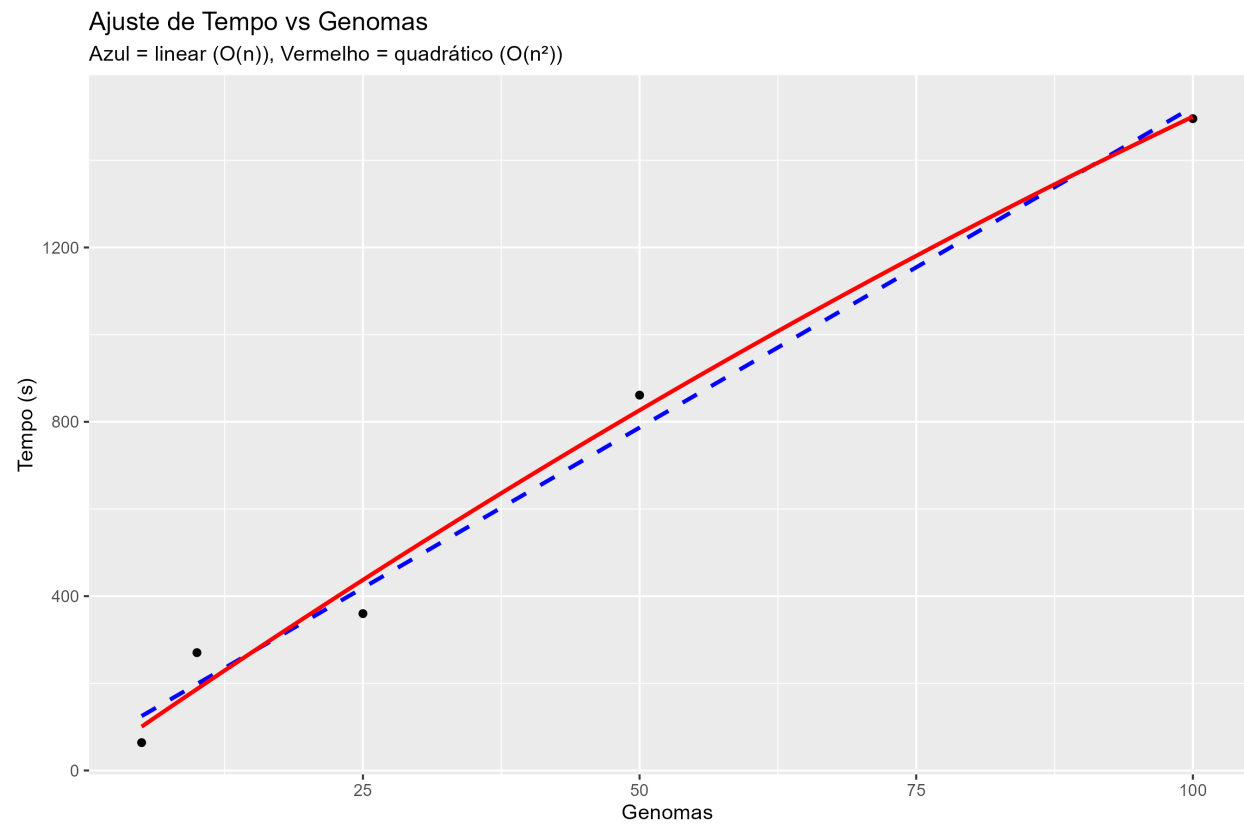
# Comparar modelos
summary(modelo_linear)
summary(modelo_quad)

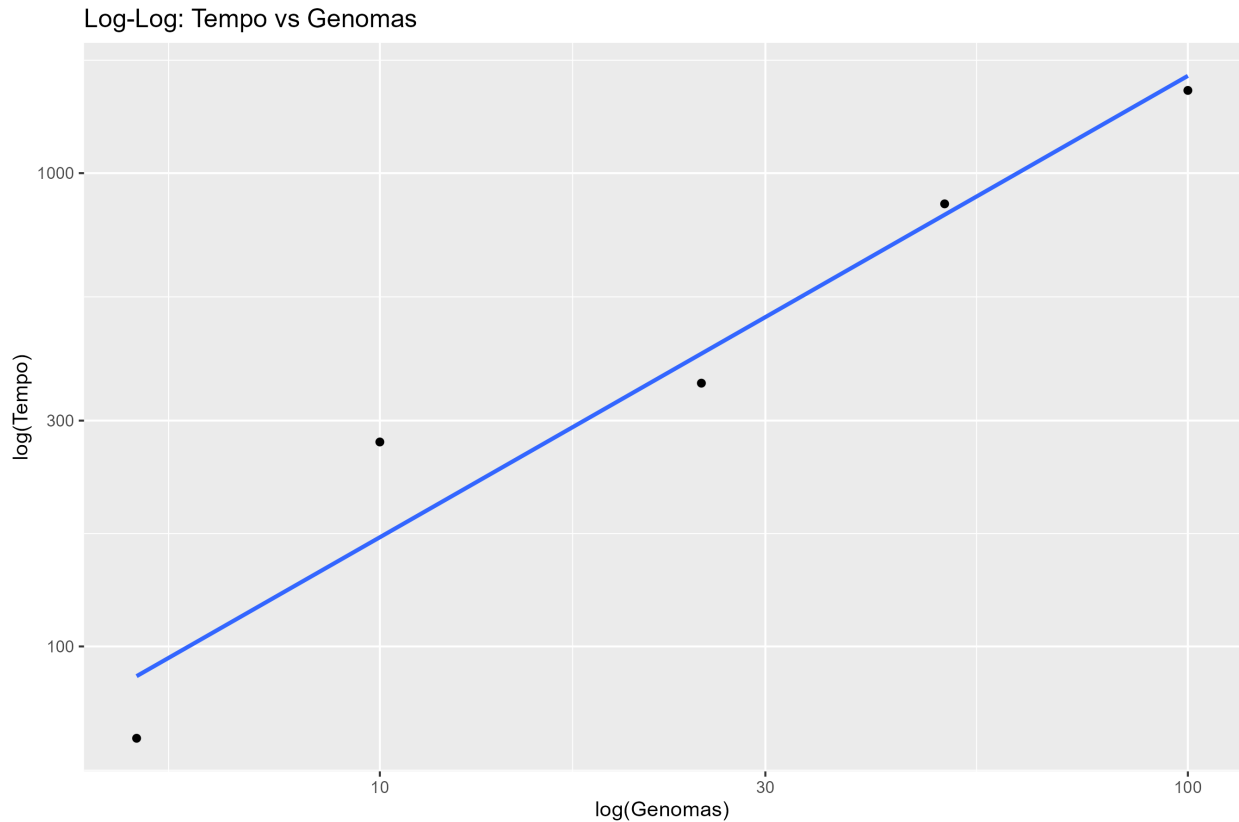
# Plotar com ajuste
p7 <- ggplot(dados_modelo, aes(x = genomas, y = elapsed_sec)) +
  geom_point() +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "blue", linetype = "dashed") +
  geom_smooth(method = "lm", formula = y ~ poly(x, 2), se = FALSE, color = "red") +
  labs(title = "Ajuste de Tempo vs Genomas",
       subtitle = "Azul = linear ( $O(n)$ ), Vermelho = quadrático ( $O(n^2)$ )",
       x = "Genomas", y = "Tempo (s)")

p7
salvar_plot(p7, "ajuste_de_tempo_vs_genomas_para_bigO")

p8 <- ggplot(dados_modelo, aes(x = genomas, y = elapsed_sec)) +
  geom_point() +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE) +
  scale_x_log10() +
  scale_y_log10() +
```

```
labs(title = "Log-Log: Tempo vs Genomas", x = "log(Genomas)", y = "log(Tempo)")
p8
salvar_plot(p8, "ajuste_de_tempo_vs_genomas_para_bigO_log_log")
```





Conclusões:

Dentro da faixa de parâmetros testados:

- Não houve grande diferença (de tempo e memória) entre a saída em `.txt` e `.txt.gz`
- Não houve grande diferença (de tempo e memória) entre a variação de `-S`

Portanto para as demais versões utilizaremos apenas:

- `-S 600`
- saída `.txt.gz`
- 25 genomas para teste de comparação de arquivos de resultados
- 100 ou mais genomas para comparação de tempo e memória

4 Versão 2.0

4.1 Modificações

- Inclusão de comentários para melhor entendimento e organização do código
- Pequenos ajustes nos textos de verbose (`[VERBOSE]`), debug(`[DEBUG]`), erro (`[ERROR]`) e aviso (`[WARNING]`) para melhor acompanhamento.
- Inclusão de limite de `depth` por segurança.


```
##/* New in Version 2.0 */
  if (depth > 1000)
    continue;
##/* --- */
```

- Limpeza ao final

```
##/* New in version 2.0 */
  labels.clear();
  rank_sp.clear();
  rank_ep.clear();
  labels.shrink_to_fit();
  rank_sp.shrink_to_fit();
  rank_ep.shrink_to_fit();
##/* --- */
```

4.2 Comparação dos arquivos de resultado

A nova versão foi compilada (make clean && make depend && make) e executada com 25 e 100 genomas. Comparado os arquivos de resultado para 25 genomas da v1.0 e v2.0

4.3 Avaliação dos logs de output (comparação de desempenho)

Tabela:

```
# Caminho dos logs
log_dir <- "v2_0/logs/output/"
log_files <- list.files(log_dir, pattern = "fsm_output_log_.*\\.txt$", full.names = TRUE)

print("Foram encontrados os arquivos")
print(log_files)

tempo_para_segundos <- function(t) {
  if (grepl(":", t)) {
    partes <- as.numeric(unlist(strsplit(t, ":")))
    if (length(partes) == 3) return(partes[1]*3600 + partes[2]*60 + partes[3])
    if (length(partes) == 2) return(partes[1]*60 + partes[2])
  }
  return(as.numeric(t))
}

extrair_dados_log <- function(arquivo) {
  linhas <- readLines(arquivo)
  nome <- basename(arquivo)

  # Extrair metadados do nome do arquivo
  versao <- str_match(nome, "fsm_output_log_(v[0-9_]+)")[,2]
  genomas <- as.numeric(str_match(nome, "_([0-9_]+)genomas")[,2])
}
```

```

max_param <- as.numeric(str_match(nome, "_([0-9]+)_max")[,2])
tipo_output <- str_match(nome, "_([TXT|GZ]--)")[,2]
timestamp <- str_match(nome, "--([0-9]{4}-[0-9]{2}-[0-9]{2}_[0-9-]{8})")[,2]

get_valor <- function(chave) {
  linha <- grep(chave, linhas, value = TRUE)
  if (length(linha) > 0) {
    return(str_trim(gsub(".*:\\s*", "", linha)))
  } else {
    return(NA)
  }
}

# Extraí tempo decorrido como string tipo "4:31.51"
get_elapsed <- function(linhas) {
  linha <- grep("Elapsed \\(wall clock\\) time", linhas, value = TRUE)
  if (length(linha) > 0) {
    tempo <- str_match(linha, ".*:\\s*(\\d+:\\d+(\\.\\d+)?)")[,2]
    return(tempo)
  } else {
    return(NA)
  }
}

# Extração dos dados do conteúdo
elapsed <- tempo_para_segundos(get_elapsed(linhas))
user <- as.numeric(get_valor("User time"))
system <- as.numeric(get_valor("System time"))
cpu_pct <- as.numeric(gsub("%", "", get_valor("Percent of CPU")))
max_mem_kb <- as.numeric(get_valor("Maximum resident set size"))
max_mem_mb <- round(max_mem_kb / 1024, 2)

#print(elapsed)

data.frame(
  arquivo = nome,
  versao = versao,
  genomas = genomas,
  max_param = max_param,
  tipo_output = tipo_output,
  timestamp = timestamp,
  elapsed_sec = elapsed,
  user_time = user,
  system_time = system,
  cpu_percent = cpu_pct,
  max_rss_mb = max_mem_mb
)
}

# Aplicar a função
tabela_logs <- bind_rows(lapply(log_files, extrair_dados_log))

# Visualizar

```

```
print(tabela_logs)
View(tabela_logs)

# Exportar se quiser
write.csv(tabela_logs, "v2_0/resumo_execucoes_v2_0.csv", row.names = FALSE)
```

Graficos: