

Optimising Existing Software with Genetic Programming

SUPPLEMENTARY INFORMATION

William B. Langdon and Mark Harman

disabled bt2_search.cpp line 1931 IF (metricsStderr)

disabled bt2_io.cpp line 612 IF (r != (ssize_t)(offsLen << 2))
replaced bt2_io.cpp line 622 for1 uint32_t i = 0 with int i = 0
replaced bt2_io.cpp line 622 for2 i < offsLenSampled with i < this->_nPat
replaced bt2_io.cpp line 622 for3 i++ with i += 3

inserted initiedRef_ = true; before aligner_sw.cpp line 145 (static_cast<void> (0));

replaced aln_sink.cpp line 758 IF (readIsPair()) with (nunpair2 > 0)
disabled aln_sink.cpp line 919 IF (nunpair1 > 0)

replaced sa_rescomb.cpp line 50 for2 i < satup_->offs.size() with 0
disabled sa_rescomb.cpp line 51 IF (satup_->offs[i] == 0xfffffffff)
deleted sa_rescomb.cpp line 52 needResolving++;
disabled sa_rescomb.cpp line 66 for2 i < refscan_.size()
deleted sa_rescomb.cpp line 68 nfound++;
disabled sa_rescomb.cpp line 69 for2 j < satup_->offs.size()
disabled sa_rescomb.cpp line 69 for3 j++
replaced sa_rescomb.cpp line 70 IF (satup_->offs[j] == refscan_[i]) with (refscan_.empty())
deleted sa_rescomb.cpp line 72 found_[i] = false;
deleted sa_rescomb.cpp line 73 nfound--;
disabled sa_rescomb.cpp line 112 IF (refscan_[i] == off)

replaced aligner_swsse_ee_u8.cpp line 707 vh = _mm_max_epu8(vh, vf); with vmax = vlo;
inserted vlo = _mm_xor_si128(vlo, vlo); before aligner_swsse_ee_u8.cpp line 711 pvHStore += 4;
replaced aligner_swsse_ee_u8.cpp line 746 ve = _mm_load_si128(pvEStore); with met.dpsucc++;
deleted aligner_swsse_ee_u8.cpp line 766 pvFStore += 4;
replaced aligner_swsse_ee_u8.cpp line 772 _mm_store_si128(pvHStore, vh); with
vh = _mm_max_epu8(vh, vf);
deleted aligner_swsse_ee_u8.cpp line 773 pvHStore += 4;
deleted aligner_swsse_ee_u8.cpp line 776 vh = _mm_subs_epu8(vh, rdgapo);
deleted aligner_swsse_ee_u8.cpp line 777 vh = _mm_subs_epu8(vh, *pvScore);
deleted aligner_swsse_ee_u8.cpp line 778 ve = _mm_max_epu8(ve, vh);
replaced aligner_swsse_ee_u8.cpp line 779 _mm_store_si128(pvEStore, ve); with
d.maxPen_ = d.maxBonus_ = 0;
deleted aligner_swsse_ee_u8.cpp line 781 pvScore += 2;
replaced aligner_swsse_ee_u8.cpp line 785 pvFStore -= colstride; with met.gathcell++;
inserted btnstack_.clear(); before aligner_swsse_ee_u8.cpp line 788 vh = _mm_load_si128(pvHStore);
deleted aligner_swsse_ee_u8.cpp line 789 pvEStore -= colstride;
replaced aligner_swsse_ee_u8.cpp line 796 vh = _mm_load_si128(pvHStore); with d.bias_ = 0;
replaced aligner_swsse_ee_u8.cpp line 801 vf = _mm_subs_epu8(vf, rfgape); with btnstack_.expand();
deleted aligner_swsse_ee_u8.cpp line 802 vf = _mm_subs_epu8(vf, *pvScore);
inserted rfgape = ((__m128i)__builtin_ia32_pshufw ((__v8hi)rfgape, 0)); before
aligner_swsse_ee_u8.cpp line 806 cmp = _mm_movemask_epi8(vtmp);
deleted aligner_swsse_ee_u8.cpp line 807 nfixup++;
disabled aligner_swsse_ee_u8.cpp line 1283 IF (!btnstack_.empty())

Fig. 1. Evolved Solution. Best of generation 200 evolved version changes 39 lines in six Bowtie2 source files.

BEST INDIVIDUAL IN GENERATION 135

Due to space constraints it is not possible to include a description of the best individual from generation 135 or its operation in the main text of our article. Therefore they are given here. See Figure 2.

The best member of generation 135 consists of 90 substitution, deletion or line insertions. Together these modify seven source files (bt2_search.cpp, bt2_idx.cpp, bt2_io.cpp, aligner_sw.cpp, aligner_result.cpp, sa_rescomb.cpp and aligner_swsse_ee_u8.cpp). Which in total take 7.25 seconds to re-compile.

Taking as an example the least effected of these, aligner_result.cpp, there are two modifications: `<_aligner_result_52>+<_aligner_result_62>` which inserts a copy of line 62 before line 52 and `<_aligner_result_1152>` which deletes line 1152. Line 62 is

```
cPretrimSoft_ = true;
```

whilst line 52 is

```
trimSoft_ = false;
```

Thus line 52 becomes `cPretrimSoft_ = true;trimSoft_ = false;`

Notice since `cPretrimSoft_` is not used between line 52 and line 62, setting it to `true` twice has no effect (except to increase the number of lines used). Whilst line 1152 is

```
left = right + diff - 1;
```

Line 1152 occurs at the end of a for loop where both `right` and `diff` may be updated. (However `right` is usually `left + 1` and `diff` is usually 1.) It appears that the usual effect of line 1152 is to step over the part of the data which has been processed by the current iteration of the loop and so not updating `left` has a safe effect of sometimes requiring work already done to be repeated. (This does not prevent loop termination, which is controlled by a separate variable.)

On its five training strands of DNA the best of generation 135 executes a total of 3 977 157 instrumented source lines. (This is 0.4% of the number of lines executed by the unmodified code). On the first two DNA sequences it finds identical answers (i.e. 0 and 1 matches respectively). On the third example Bowtie2 finds two poor matches, mean quality 26.02, i.e. on average almost 10 worse than a perfect match, whereas the generation 135 code reports zero matches. On the last two test cases it finds 3 matches (versus 30) and 10 matches (versus 116). However the average quality of matches on the last two test sequences is higher (30.89 v. 29.20 and 34.60 v. 33.88).

```

<IF_aligner_sw_1193> <IF_bt2_search_1954> <_bt2_idx_1334> <_aligner_sw_252>
<IF_aligner_sw_1201><IF_aligner_sw_1193> <IF_aligner_sw_251><IF_aligner_sw_101>
<IF_aligner_swsse_ee_u8_1351> <_aligner_swsse_ee_u8_746>+<_aligner_swsse_ee_u8_745>
<IF_bt2_search_1995> <_aligner_swsse_ee_u8_805>

<_aligner_swsse_ee_u8_717><_aligner_swsse_ee_u8_342>
<_aligner_swsse_ee_u8_756><_aligner_swsse_ee_u8_359> <_aligner_swsse_ee_u8_806>
<_bt2_search_1985><_bt2_search_1662> <_aligner_swsse_ee_u8_739>
<_aligner_swsse_ee_u8_745> <_aligner_swsse_ee_u8_762><_aligner_swsse_ee_u8_807>
<_aligner_swsse_ee_u8_773> <_aligner_result_52>+<_aligner_result_62>
<_aligner_swsse_ee_u8_360>

<_aligner_swsse_ee_u8_733> <_aligner_swsse_ee_u8_1338>
<for1_sa_rescomb_69><for1_sa_rescomb_50> <_aligner_swsse_ee_u8_755>
<IF_bt2_search_1636> <for3_sa_rescomb_66> <_bt2_search_1426> <for2_sa_rescomb_66>
<for3_aligner_sw_402> <_aligner_swsse_ee_u8_776><_aligner_swsse_ee_u8_343>

<_aligner_swsse_ee_u8_804> <_aligner_swsse_ee_u8_765>
<IF_bt2_idx_1399><IF_bt2_idx_1403> <_sa_rescomb_52><_sa_rescomb_64>
<_aligner_swsse_ee_u8_807><_aligner_swsse_ee_u8_361>
<for2_sa_rescomb_69><for2_sa_rescomb_50> <_sa_rescomb_67>+<_sa_rescomb_44>
<_bt2_idx_1343> <IF_aligner_sw_300><IF_aligner_sw_2040>
<IF_aligner_swsse_ee_u8_1235><IF_aligner_swsse_ee_u8_1101>

<_aligner_swsse_ee_u8_766> <_aligner_swsse_ee_u8_743>
<for3_sa_rescomb_69><for3_sa_rescomb_50> <_aligner_swsse_ee_u8_740>
<_aligner_swsse_ee_u8_394> <_aligner_swsse_ee_u8_779> <_bt2_io_623><_bt2_io_517>
<_aligner_swsse_ee_u8_726><_aligner_swsse_ee_u8_957> <_aligner_swsse_ee_u8_780>
<_aligner_swsse_ee_u8_796>

<for1_bt2_io_622><for1_bt2_io_421> <for2_sa_rescomb_50><for2_sa_rescomb_66>
<_aligner_swsse_ee_u8_769><_aligner_swsse_ee_u8_680>
<_bt2_search_1985>+<_bt2_search_1843> <for2_bt2_io_622><for2_bt2_io_448>
<_sa_rescomb_68>+<_sa_rescomb_73> <_bt2_idx_138> <_aligner_swsse_ee_u8_778>
<for3_bt2_io_622><for3_bt2_io_321> <_sa_rescomb_73>+<_sa_rescomb_64>

<_aligner_swsse_ee_u8_797><_aligner_swsse_ee_u8_393>
<IF_aligner_swsse_ee_u8_784> <_sa_rescomb_72> <IF_aligner_swsse_ee_u8_1283>
<IF_sa_rescomb_70> <_aligner_swsse_ee_u8_1074>
<_aligner_swsse_ee_u8_727>+<_aligner_swsse_ee_u8_947>
<_aligner_swsse_ee_u8_801>+<_aligner_swsse_ee_u8_363>
<_aligner_swsse_ee_u8_802> <_aligner_swsse_ee_u8_772>

<_aligner_swsse_ee_u8_803><_aligner_swsse_ee_u8_365>
<_sa_rescomb_71>+<_sa_rescomb_43> <for2_aligner_sw_233><for2_aligner_sw_1188>
<_aligner_result_1152> <for3_sa_rescomb_50><for3_sa_rescomb_69>
<_aligner_swsse_ee_u8_795>+<_aligner_swsse_ee_u8_364>
<for3_sa_rescomb_111><for3_sa_rescomb_69>
<_aligner_swsse_ee_u8_695><_aligner_swsse_ee_u8_1074> <IF_sa_rescomb_51>
<_aligner_swsse_ee_u8_777>

<_aligner_swsse_ee_u8_755>+<_aligner_swsse_ee_u8_740>
<_aligner_swsse_ee_u8_781><_aligner_swsse_ee_u8_955>
<for2_sa_rescomb_111><for2_sa_rescomb_50> <for1_sa_rescomb_50><for1_sa_rescomb_69>
<for1_sa_rescomb_66><for1_sa_rescomb_69> <IF_sa_rescomb_112> <_bt2_idx_920>
<for1_sa_rescomb_111><for1_sa_rescomb_69> <_bt2_search_3095><_bt2_search_278>
<_aligner_swsse_ee_u8_793><_aligner_swsse_ee_u8_958>

```

Fig. 2. Best member in generation 135