Optimising Existing Software with Genetic Programming
SUPPLEMENTARY INFORMATION

William B. Langdon and Mark Harman
Fig. 1. Evolved Solution. Best of generation 200 evolved version changes 39 lines in six Bowtie2 source files.
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_eu8.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).
Fig. 2. Best member in generation 135