Iteration	Purpose	Defined	Due	Delivered	Version	Notes
1	Initial with basic infrastructure	19/02/2010	12/03/2010	12/03/2010	1.00	
2	File I/O and data structures	12/03/2010	01/04/2010	06/04/2010	1.01	
3	AYB Initialisation	08/04/2010	26/04/2010	27/04/2010	1.02	
4	Base Call Loop	06/05/2010	21/05/2010	20/05/2010	1.03	
5	MPN, quality output, datablock	28/05/2010	17/06/2010	17/06/2010	1.04	
6	P Solver, MPN unit test, cif format Pause for testing and experimentation	22/06/2010	29/07/2010	30/07/2010	1.05	
7	Final working values, mu arg, bad data handling	25/08/2010	21/09/2010	22/09/2010	1.06	
8	Quality score calibration, missing data handling, GC content, sim output, argument defaults, optimisation.	01/11/2010	12/11/2010	15/11/2010	1.07	

Index	Priority	Task	Planned	Actual	Notes
1		Initial system: Make; Run	1	1	
2		Program Version	1	1	
3		Program arguments: Read and store: Infrastructure	1	1	
4		Program Log: Infrastructure	1	1	
5		Signal Handler	1	1	
6		File I/O: Locate, open/create, read/write, close	2	2	Requires args input (i), output (o);
7		Data Structures: Tile, Cluster, Matrix	2	2	
8		Intensities Input: Locate, read, store, tidy up	2	2	
9		Matrix Input: Locate, read, store, tidy up; Read M, N, P	2	3	Requires args M (M), N (N), P (P)
10		Processed Intensities: Calculate and store, tidy up	3	3	
11		Initial Sequence: By maximum intensity	3	3	
12		Sequence output: Create file, write data, close	3	3	
13		Replace message macros with vfprintf	2	2	
14		Expand message severity list to include debug	2	2	
15		Configure Doxygen; add comments to new HM files	2	2	
16		Matrix: Add read_methods for multiple styles	3	3	1) Intensities 2) As rows of columns. Move matrix read from cluster.
17		Initial Lambda: Ignore weights	3	3	

18	Calculate Covariance	4	4	
19	Estimate Lambda	4	4	
20	Call Bases	4	4	
21	Base Call Loop	4	4	Requires args niter (n?); change ncycles to c?
22	Utility: Reimplement xfree null return	4	4	Use ** or return, could apply to all free_* functions
23	XIO: Check and amend file structure handling; nulls etc.; null return on close	4	4	Make interface as close to normal file handle as possible
24	XIO: rename initialise_aybstd to not contain 'ayb'	4	4	
25	I/O: If input file open fails then warn and carry on	4	4	But stop if output fails as all will
26	I/O: Output files should not be compressed	4	4	Test: compressed input, input with no extension, input with no delimiter, combo?
27	I/O: Create output directories if do not exist	4	4	Include log; abort if create fails.
28	I/O: Intensities: Expand fixed match to “_int.txt”	4	4	As original
29	Program log: Replace fixed “ayb” with prefix	4	4	hhmm not sufficient for parallel runs
30	Program log: Add initial information line	4	4	Program name, date/time
31	Program log: switch order of warning and information	4	4	

32	Process Intensities file for analysis by block	5	5	New argument blockstring (b) of the form: RnInCn, decoded as: R=>read I=>ignore C=>concatenate to previous block Note no difference in analysis for forward and backward data. Then: less data than specified=>abort program; more data than specified=>warn and continue
33	Implement Quality Scores	5	5	Actually 2 alternative outputs fasta/fastq. Requires arg format (f).
34	Implement MPN estimation.	5	5	Get from AYBc; convert intensities access from single array to via cluster list. Use MPN initialisation as in AYBc
35	Allow selection of routines to solve for P. Remove simultaneous solve for P and N (from TM)	6	6	Alternatives are: Standard SVD; Standard SVD then zero negative entries; Non-negative least squares.
36	Unit test of MPN estimation.	6	6	Already exists in an old form but requires changes; coercing values into array/cluster/tile; other bits.
37	Option to read Intensity data in cif format.	6	6	Get from AYBc; convert intensities access from single array to via cluster list. Requires arg dataformat (d).
38	Optional output of final processed intensities, M, P, N, lambda, weights	7	7	Final processed format as intensities input, illlumina or cif. Write cif format already exists. One file per input with tag “pif”. Single file (tag “final”) used for all other values in show_MAT format. Requires arg working (w)
39	mu should be a parameter as in AYBc.	7	7	Requires arg mu (m)
40	Introduce a diagonal delta to the solver routines (ridge regression); avoids failure to solve when data bad .	7	7	Use value 1 which is small compared with typical matrix values.
41	Count zero lambdas for each file and output at end if any.	7	7	Output per iteration.
42	Stop and issue message if data error detected.	7	7	Initially failure to calculate covariance.
43	Remove padding zeros from output cluster line.	7	7	There are typically > 10,000 clusters but could be any number.
44	Couple of bugs in string length allocation.	7	7	Cause segment errors.

45	Make double calculations standard.	7	7	Accuracy of float not sufficient for larger values.
46	Calibration of quality score.	8	8	Adjust in line with empirical observations using a table and neighbouring scores.
47	Deal with missing data by setting base call to ambiguous.	8	8	Defined as all intensities zero.
48	Additional deltas; use for M solve and add for covariance.	8	8	
49	Option to penalise base calls by genomic CG content.	8	8	Requires arg composition (c); values 0<gc<1.
50	Tidy compiler warnings from fortran and array.def.	8	8	
51	Optional output of full covariance matrix and lambda fit.	8	8	For multiple blocks use lambdas from first block only and append additional block covariance. Print header text, program version and command line, substituting for header text as already printed, then num cycles and fit params. All lines to begin with hash; locate any newlines and add hash. Requires arg simdata (s)
52	Final processed output does not allow for multiple blocks.	8	8	
53	Changes to arguments: default dataformat to cif, format to fastq, solver to zero. New blockstring default all available cycles in one block. Replace prefix with any number of non-option arguments, also allowing inclusion of partial path.	8	8	Requires additional input file search loop. Use of prefix has caused problems when e.g. file and file_end1 both exists. Make default prefix behaviour an exact match with a '+' as last character indicating treat as prefix. Message file must now be supplied as path and filename.
	Flush error file after each message to capture info in case of failure.	8	8	with 53
	Change message to use CSTRING instead of fixed length	8	8	with 53
54	Ensure an info message for all options that affect results.	8	8	Need GC comp and Mu; ignore message file. No genome composition if default.
55	Optimisation	8	8	Improve runtime.

56	Tidy compiler warnings from intel C compiler.	8	8 strcasecmp in strings, mode_t in sys/types, BSD_SOURCE for scandir, cast int to enum, goto jumps over variable declarations, multiple const
57	Additional output of M, P and N matrices in correct format for input with final processed. Does not seem to respond to ctrl<c>		
	Tidy function headers for doxygen.		Weibull

New

	Method of filtering the list of clusters according to some criteria; resultant finds should be marked (separately)	Possible filters: - cycle in cluster with all signals zero (signals missing) - cycle in cluster with signal greater than some threshold (argument) The filter_list, filtercopy_list and split_list functions do this but may need to be modified to retain the same ordering of elements (cf read_TILE and read_known_TILE).
	Look at efficiency.	
	Automate module testing with a test script. Create reference results for future comparison.	Enhancements to cluster/tile; null values, cif.

Errors

In Matrix, test instructions are wrong way round	
Functions that take no arguments should include void inside the argument brackets.	Deal with as seen. main tidyup, options init_options, handler checksignals.

Consider

Check all algorithms use preferred versions.	e.g. calculate lambda.
AYB struct: Make bases and quals part of cluster, also lambda and weights?	But will make tile bigger with empty space? weights are accessed as an array for mean/var purpose; also use of set/scale_MAT.
Parameters to makefile to allow debug/double?	

Refactor

Split ayb_model	Possibly extract all AYB into new ayb. It's possible that static variables are inefficient - may want to pass AYB pointer in but keep internals hidden.
Move base penalty to call_base.	Where it is used.
Rationalise use of int/long/uint32 types	Be aware of efficiency issues.
Optionally use message in matrix/cluster/tile/nuc/more? controlled by compiler switch.	Set switch flag in message?
New message functionality to return message string	For use with perror, length of string? vsnprintf?
Message: use a different word than type for msgtype?	Confusing in documentation
Add segmentation violation handler	
Tidy up order of some items, e.g. ayb_model set.., list of options.	

Make --version option comply to GNU coding standards

Consider

ayb_model: make bases & quals a list (per cluster)

ayb_model: does AYB struct need to be public at all? or even the standard functions?

xio, matrix, cluster, tile not yet documented

Index	Note	Resolution
1	Divide by zero does not cause FPE so not tested	
2	Have not resolved all the get memory issues during data append.	

Index	Task	Additional Requirements	Completed	Notes
46	Calibration of quality score.	Adjust in line with empirical observations using a table and neighbouring scores.	12/10/2010	Call base returns quality value not phred char; new adjust quality called from ayb_model after all bases called and before convert to phred; first and last require special handling; include new calibration files. New nuc routines to split prob to phred and MIN_QUALITY.
47	Deal with missing data by setting base call to ambiguous.	Defined as all intensities zero.	12/10/2010	Nuc ambig is beyond nbase array range; new nuc isambig called wherever base index referenced. New call_base_nodata; add handling to quality adjust. In ayb_model check for no data before call base.
48	Additional deltas; use for M solve and add for covariance.		12/10/2010	
49	Option to penalise base calls by genomic CG content.	Requires arg composition (c); values 0<gc<1.	12/10/2010	Penalty calculated for each base and stored as arrayx4. Pass to call base and use to adjust calculation.
50	Tidy compiler warnings from fortran and array.def.		12/10/2010	Array use attribute used.
51	Optional output of full covariance matrix and lambda fit.	For multiple blocks use lambdas from first block only and append additional block covariance. Print header text, program version and command line, substituting for header text as already printed, then num cycles and fit params. All lines to begin with hash; locate any newlines and add hash. Requires arg simdata (s)	22/10/2010	Dirio open append option. Matrix show_MAT with optional rownum. Ayb_options match string to an option. Version now in new ayb_version.c so version can be exported to more than one file. Ayb_model store flag and header text; accumulate_all_covariance; output_simdata calculates values and writes after formatting header lines; analyse_tile needs arg params from ayb_main.
52	Final processed output does not allow for multiple blocks.		22/10/2010	ayb_model pass blk ->estimate_bases -> open_processed; use to open pif and final.

53a	Changes to arguments: default dataformat to cif, format to fastq, solver to zero.		27/10/2010	dataformat - dirio change default; ayb_model remove duplicate InputFormat - get when needed; info message moved to dirio (improve order). output format - ayb_model change default. solver - ayb_model change default; store solver index and output info message; make selection info message generic. Update help defaults.
53b	New blockstring default all available cycles in one block.		28/10/2010	
53c	Replace prefix with any number of non-option arguments, also allowing inclusion of partial path.	Requires additional input file search loop. Use of prefix has caused problems when e.g. file and file_end1 both exists. Make default prefix behaviour an exact match with a '+' as last character indicating treat as prefix. Message file must now be supplied as path and filename.	09/11/2010	Dirio move pattern check/scandir to set_pattern; new clear_pattern; new pattern path, filled by new move_partial_path, to hold input path and partial path; in match_pattern treat as prefix only if prefix indicator. Read_options return next arg index; options/help/usage make alphabetical. Move retopt enum to utility and use as analyse_tile return to indicate stop/next pattern; propagate in ayb_model as required. Ayb_main new outer loop through non-option args calling set_pattern. Message path is now full pathname not location and no default; create_filename replaced by check_message_path; some fatal messages now error.
54	Ensure an info message for all options that affect results.	Need GC comp and Mu; ignore message file. No genome composition if default.	02/11/2010	Ayb_model output from model startup; mu needs export from call_bases.
55	Optimisation	Improve runtime.	11/11/2010	Nuc new has_ambiguous_base used as precheck where isambig is called in tight loop; replace int/uint_32 with uint_fast32 in compute heavy loops; precalculate weight matrix; affects intensities process/expected_intensities, mpn calculateJ/K. Makefile add unroll-loops optimiser.
56	Tidy compiler warnings from intel C compiler.	strcasecmp in strings, mode_t in sys/types, BSD_SOURCE for scandir, cast int to enum, goto jumps over variable declarations, multiple const	12/11/2010	ayb_main, ayb_model, cif, cluster, dirio, tile, utility, Makefile

Index	Task	Additional Requirements	Completed	Notes
38	Optional output of final processed intensities, M, P, N, lambda, weights	Final processed format as intensities input, illlumina or cif. Write cif format already exists. One file per input with tag “pif”. Single file (tag “final”) used for all other values in show_MAT format. Requires arg working (w)	20/09/2010	Matrix write_to_line; Cluster write_coords; new read_first to get lane & tile, common read parts to subproc; Tile write_lane_tile, use cluster read_first. Cif set_from_real, create_cif, option to show all; Utility new real_t round function definition. Ayb_model standard (write each line) or cif (store and write at end) output; open/write/close_processed; pass last iter flag to estimate_bases; option to show only part AYB structure; new input format info message.
39	mu should be a parameter as in AYBc.	Requires arg mu (m)	21/09/2010	Adjusts range of quality scores; smaller value produces higher maximum quality score.
40	Introduce a diagonal delta to the solver routines (ridge regression); avoids failure to solve when data bad .	Use value 1 which is small compared with typical matrix values.	26/08/2010	Value stored in ayb_model as constant DELTA_DIAG.
41	Count zero lambdas for each file and output at end if any.	Output per iteration.	25/08/2010	Store in ayb_model. Create string for message with count for each iteration.
42	Stop and issue message if data error detected.	Initially failure to calculate covariance.	26/08/2010	New call_base_null, same as used for zero lambda. estimate_bases can return err val.
43	Remove padding zeros from output cluster line.	There are typically > 10,000 clusters but could be any number.	21/09/2010	In ayb_model output_results.
44	Couple of bugs in string length allocation.	Cause segment errors.	25/08/2010	In dirio: output_name(_cif), calculation of newname length requires bracket around blk query part or result seems to always be 1; scan_inputs fixed buffer insufficient.
45	Make double calculations standard.	Accuracy of float not sufficient for larger values.	25/08/2010	Rename original makefile to makefloat (not controlled).

Index	Task	Additional Requirements	Completed	Notes
35	Allow selection of routines to solve for P. Remove simultaneous solve for P and N (from TM).	Alternatives are: Standard SVD; Standard SVD then zero negative entries; Non-negative least squares. New arg solver (S).	09/07/2010 28/07/2010	Changes to help, model, mpn, matrix error. Need fortran compiler switch out (NFORTRAN) for eclipse.
36	Unit test of MPN estimation.	Already exists in an old form but requires changes; coercing values into array/cluster/tile; other bits.	23/07/2010	coerce matrix and expanded to array/cluster/tile. Calculate intermediates and new values.
37	Option to read Intensity data in cif format.	Get from AYBc; convert intensities access from single array to via cluster list. Requires arg dataformat (d).	21/07/2010	Adopt cif. Store input format in dirio; alternatives for file search and output filename. New read_cif_tile/cluster. In model alternative read and output filename. New debug output constraints.

Index	Task	Additional Requirements	Completed	Notes
32	Process Intensities file for analysis by block	New argument blockstring (b) of the form: InRnCn, decoded as: I=>ignore R=>read C=>concatenate to previous block Note no difference in analysis for forward and backward data. Then: less data than specified=>abort program; more data than specified=>warn and continue Add letter extension to output file name if more than one block.	09/06/2010	Replace option cycles (c) with blockstring (b). New datablock class for datablock structure. Create tile/cluster/matrix append functions. Pre-process input tile and create an array of sub-tile pointers. Add ncycle to TILE structure. No longer need ncycle return from tile. Change action if not enough cycles; message if spare cycles (pass flag on to read matrix line to check if first line). In read_tile error if later clusters have < cycles than previous; indicates a faulty file. New open_output_blk adds a block suffix letter. Analysis now a loop.
33	Implement Quality Scores	Actually 2 alternative outputs fasta/fastq. Requires arg format (f).	16/06/2010	Function match_string to utility.
34	Implement MPN estimation.	Get from AYBc; convert intensities access from single array to via cluster list. Use MPN initialisation as in AYBc.	15/06/2010	Need new modules mpn, statistics; new functions update_cluster_weights, estimate_MPN (model), expected_intensities (intensities), some matrix and lapack. Read-in matrices become optional; new dirio func to say matrix specified.

Index	Task	Additional Requirements	Completed	Notes
18	Calculate Covariance		11/05/2010	Put in ayb_model not call_bases because uses structure of AYB. Scale reciprocal removed.
19	Estimate Lambda		12/05/2010	Use estimate_lambdaWLS as originally described.
20	Call Bases		11/05/2010	Constant Mu used for quality score.
21	Base Call Loop	Requires args niter (n); change ncycles to c	12/05/2010	Niter static in ayb_model.
22	Utility: Reimplement xfree null return	Use ** or return, could apply to all free_* functions	13/05/2010	Used by free_AYB/CSTRING/(MAT)/CLUSTER/TILE. Leave xfree returning void and instead return null pointer from free_functions. Do not have to have a return in call to free_x if not needed, e.g. a local var. ARRAY/LIST never freed where return matters.
23	XIO: Check and amend file structure handling; nulls etc.; null return on close	Make interface as close to normal file handle as possible	17/05/2010	Free structure and return null pointer if open fails. Return null pointer on close (different from normal file handle operation)
24	XIO: rename initialise_aybstd to not contain 'ayb'		13/05/2010	
25	I/O: If input file open fails then warn and carry on	But stop if output fails as all will	17/05/2010	Dirio open_output: loop until successful open.
26	I/O: Output files should not be compressed	Test: compressed input, input with no extension, input with no delimiter, combo?	18/05/2010	
27	I/O: Create output directories if do not exist	Include log; abort if create fails.	18/05/2010	Exist, no exist, exist file not dir.
28	I/O: Intensities: Expand fixed match to “_int.txt”	As original	17/05/2010	
29	Program log: Replace fixed “ayb” with prefix	hhmm not sufficient for parallel runs	19/05/2010	
30	Program log: Add initial information line	Program name, date/time	19/05/2010	AYB Message Log; user name and datetime
31	Program log: switch order of warning and information		19/05/2010	

Index	Task	Additional Requirements	Completed	Notes
9	Matrix Input: Locate, read, store, tidy up; Read M, N, P	Requires args M (M), N (N), P (P);	13/04/2010 16/04/2010	Read and written, not stored. Stored in new ayb_model.
10	Processed Intensities: Calculate and store, tidy up	AYB struct taken from AYBc with int16 intensities replaced with tile	20/04/2010	
11	Initial Sequence: By maximum intensity	use call_base_simple	20/04/2010	
12	Sequence output: Create file, write data, close		21/04/2010	NUC needs raw type file? NUC changed to use XFILE
16	Matrix: Add read_methods for multiple styles	1) Intensities 2) As rows of columns. Move matrix read from cluster.	13/04/2010	New functions new_MAT_from_line, read_MAT_from_column_file
17	Initial Lambda: Ignore weights	use estimate_lambdaOLS	26/04/2010	

Notes

array.def	Changes from git central	New rcons*_list to appends to a given node (should be last)
tile	Changes from git central	New read_TILE keeps cluster list in input file order
dirio	Store location of predetermined matrices; new open_matrix; new method returns name of current file (for message).	set_path becomes more generic set_location;
matrix	set_MAT, transpose_inplace, invert taken from AYBc.	In invert change WORK/WORKSPACE type to real_t
lapack	For matrix, getrf/i taken from AYBc.	change WORK type to float in sgetri
nuc	Taken from AYBc; defines NUC and PHREDCHAR types; isprob taken from utility.	Read/show changed to use XFILE; show_PHREDCHAR print space if out of range; replace printf with message NUC_*, *_PHRED do not need to be public? NUC_ to enum? What about array construct?

call_bases	New; parts taken from AYBc call_bases (more later).	
intensities	New; parts taken from AYBc process intensities (more later).	
lambda	New; parts taken from AYBc estimate_lambda (more later).	
ayb_model	New; parts taken from AYBc ayb. (more later).	show_AYB use fp not stderr
utility.h	Remove NBASE def - now in nuc	

Consider

ayb_model	Make bases & quals a list (per cluster)	
ayb_model	Does AYB struct need to be public at all? or even the standard functions?	

Index	Task	Additional Requirements	Completed	Notes
6	File I/O: Locate, open/ create, read/write, close	Requires args input (i), output (o); search input dir for pattern matched files.	24/03/2010	
7	Data Structures: Tile, Cluster, Matrix	Use as is from Central Repository	01/04/2010	
8	Intensities Input: Locate, read, store, tidy up	Requires arg ncycles (n)	01/04/2010	
9	Matrix Input: Locate, read, store, tidy up; Read M, N, P	Use ncycle	postponed	
13	Replace message macros with vfprintf	Change call function name to message	12/03/2010	
14	Expand message severity list to include debug		24/03/2010	
15	Configure Doxygen; add comments to new HM files		29/03/2010	Adopted files still to do

Index	Task	Additional Requirements	Completed	Notes
1	Initial system: Make; Run	IDE develop but also build from command line	10/03/2010	Make with gcc. Makefile.
2	Program Version	Version file	08/03/2010	Store version and date.
3	Program arguments: Read and store (infrastructure)	Initial args help, licence, version, usage (default)	08/03/2010	Use getopt_long; Use include file method to do bulk output.
4	Program Log: Infrastructure	Log message from a Type and Severity; allow for parameters of varying type. Output to unique filename (from date/time) in configurable location.	10/03/2010	Requires args logfile (e), loglevel (l); Hide implementation from user.
5	Signal Handler	Initially interrupt and floating point exception	10/03/2010	On interrupt get confirmation first. Divide by zero does not cause FPE so not tested