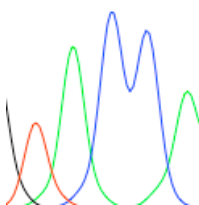## PEAKTRACE™ BASECALLER

**PeakTrace™** is a novel basecaller for improving the quality and read length of Sanger DNA sequencing traces. The PeakTrace basecaller works with trace files produced by the ABI 310, 3700, 3100, 3130, 3730, and 3500 DNA sequencers. The MegBACE™ sequencers are also supported.

PeakTrace can improve the number of high quality bases by up to 50% as compared to that obtained from the ABI KB™ basecaller. PeakTrace offers traces that are better resolved with fewer errors than any other available basecaller.

PeakTrace is available in three formats. **PeakTrace:Box™** is a standalone hardware system designed for large to medium facilities. **Auto PeakTrace RP™** is a user installable version of PeakTrace for small to medium scale users. **PeakTrace FTP™** is a online based service for small scale users.
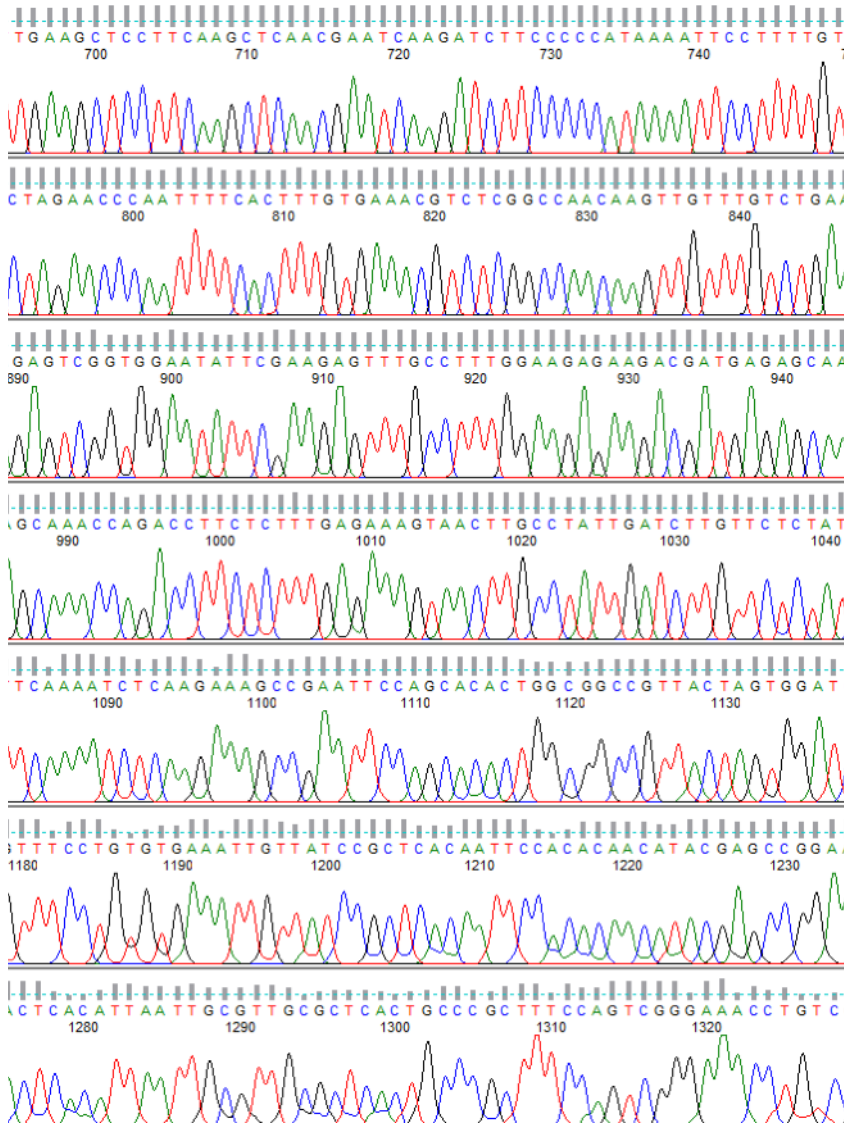
PEAKTRACE

SIMPLY BETTER BASECALLING

PeakTrace™ offers superior basecalling to existing Sanger sequencing basecallers like KB™, TraceTuner or phred.

PeakTrace can be used via a simple to use graphical interface (Auto PeakTrace), or simply integrated into existing LIMS infrastructure using the PeakTrace command line version. Users have full control of the various basecalling parameters such as output format, trace processing and/or trimming.

PeakTrace can output trace files in both .ab1 and .scf format as well as text based files such as .seq (FASTA or ABI), .phd.1 and .qual format.

PeakTrace provides a log of the basecalling results for each trace file including the basecalling result and/or the reasons for any basecalling failures.

The PeakTrace:Box™ system is available in a range of monthly volumes with additional discounts available for large scale or long term leases. Auto PeakTrace RP™ and PeakTrace FTP™ are charged per trace basecall.

PEAKTRACE:BOX
A standalone PeakTrace basecalling system designed for integration into medium to large sequencing facilities. Does not require an internet connection.
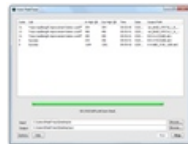
AUTO PEAKTRACE RP
User installable version of the Auto PeakTrace software where the data processing is performed on the PeakTrace server.

PEAKTRACE FTP
Online service offering PeakTrace basecalling via FTP with a guaranteed improvement. Suitable for small scale users.

VALIDATION OF PEAKTRACE BASECALLING

The use of improved basecallers offers a simple method for increasing trace read length. Before a new basecaller can be used in production, it needs to be validated to ensure that the predicted quality scores match the actual or observed quality scores. Such a process is known as quality score mapping.

PeakTrace is an alternative basecaller to the current "gold standard" basecaller for Sanger sequencing, the KB™ basecaller. The aim of this study was to validate the PeakTrace basecaller and compare the quality score mapping to that obtained using the KB basecaller.

METHODOLOGY

Two thousand, two hundred and three traces collected from pUC sub clones derived from two 175 kB Salmon BAC clones were basecalled using either KB v1.2 (ABI) or the PeakTrace basecaller (Nucleics). The traces were pre-screened using the **QualTrace™ II QC software** (Nucleics) to remove failed and mixed signal traces. To ensure that only like-against-like traces were compared, only traces where PeakTrace provided more than 10 Q20+ bases were included. All traces were BLAST (NCBI) aligned to the respective BAC consensus sequence. Traces that could not be aligned to the BAC sequenced were excluded from further analysis (the majority of these unalignable traces were derived from *E. coli* K12 genomic DNA). Finally, all reads that contained putative Q40+ errors were examined manually to remove sequences that contained obvious BLAST misalignments. A total of 1643 traces passed these four screening criteria.

The BLAST aligned sequences were then used to calculate the total aligned bases and observed errors using the approach of Ewing and Green (1). In brief, for every aligned base the total count of correct and incorrect basecalls was recorded (observed quality or Q values). These were compared to each basecaller's predicted error rates (predicted Q scores) to determine the accuracy of the quality score prediction (Q score mapping) of each basecaller.

RESULTS

The results of the Q mapping are shown in Table 1 and 2. PeakTrace was more accurate in predicting the actual error rate than KB (Figure 1). This accuracy was particular noticeable in the mid-range quality scores (Q20 to Q30) where KB significantly over predicted the trace quality.

The average total aligned read length for PeakTrace basecalled traces was 1030.4 bases. The average aligned Q20+ read length for PeakTrace basecalled traces was 954.4 bases. This compared to an average total aligned read length for KB basecalled traces (using the same 1643 traces) of 867.9 bases, with an average Q20+ read length of 795.9 bases. **PeakTrace increases the total aligned read length of this trace data set by 18.7% and Q20+ read length by 19.9%.**

Table 1. Quality score mapping of PeakTrace™ basecalled traces

QScore	Count	Errors	Error Rate	QScore	Count	Errors	Error Rate
1	1	0	-	29	714	6	20.8
2	21	15	1.5	30	43956	51	29.4
3	80	19	6.2	31	997	0	-
4	1027	165	7.9	32	5759	11	27.2
5	2400	439	7.4	33	1382	0	-
6	5296	1099	6.8	34	1230	1	30.9
7	3138	348	9.6	35	21900	5	36.4
8	33708	6606	7.1	36	2398	1	33.8
9	11406	1650	8.4	37	6001	1	37.8
10	8204	906	9.6	38	3324	2	32.2
11	11496	1115	10.1	39	1288	0	-
12	10699	751	11.5	40	41047	6	38.4
13	15127	947	12.0	41	5233	0	-
14	8423	424	13.0	42	1504	0	-
15	5419	211	14.1	43	5721	0	-
16	2956	74	16.0	44	1972	0	-
17	1783	45	16.0	45	42940	1	46.3
18	3019	48	18.0	46	289	0	-
19	615	7	19.4	47	3692	1	35.7
20	44789	588	18.8	48	149	0	-
21	4489	38	20.7	49	4317	0	-
22	9383	67	21.5	50	84302	0	-
23	7965	40	23.0	51	259	0	-
24	3429	18	22.8	52	479835	1	56.8
25	17882	67	24.3	53	973	0	-
26	494	2	23.9	54	1073	0	-
27	2292	3	28.8	55	714262	0	-
28	903	2	26.5	**Q20+**	**1568143**	**912**	
				Total	**1692961**	**15781**	

Table 2. Quality score mapping of KB™ basecalled traces

QScore	Count	Errors	Error Rate	QScore	Count	Errors	Error Rate
1	332	260	1.1	30	15823	44	25.6
2	533	141	5.8	31	9285	15	27.9
3	3815	974	5.9	32	8860	18	26.9
4	8190	1739	6.7	33	8686	20	26.4
5	7151	1431	7.0	34	7179	4	32.5
6	8951	1424	8.0	35	5431	7	28.9
7	8830	1286	8.4	36	12957	10	31.1
8	7725	959	9.1	37	8234	4	33.1
9	6787	850	9.0	38	11122	3	35.7
10	7629	662	10.6	39	9400	2	36.7
11	6450	512	11.0	40	7524	2	35.8
12	6878	425	12.1	41	12751	6	33.3
13	6899	414	12.2	42	4541	0	-
14	6503	276	13.7	43	16097	2	39.1
15	5330	171	14.9	44	10576	0	-
16	6404	163	15.9	45	9207	0	-
17	6407	145	16.5	46	23183	2	40.6
18	6807	130	17.2	47	21292	3	38.5
19	6648	98	18.3	48	4343	0	-
20	6793	88	18.9	49	5756	0	-
21	6608	61	20.3	50	8886	0	-
22	5906	63	19.7	51	15279	0	-
23	7303	65	20.5	52	22520	0	-
24	7353	46	22.0	53	47587	0	-
25	6526	50	21.2	55	18533	0	-
26	6751	50	21.3	57	99393	0	-
27	6522	33	23.0	59	69445	1	48.4
28	9110	70	21.1	61	726458	0	-
29	24515	102	23.8	**Q20+**	**1307735**	**771**	
				Total	**1426004**	**12831**	

CONCLUSION

PeakTrace offers improve read lengths of 20% in alignable Q20+ bases over that obtained with the KB basecaller. PeakTrace can be used with trace data that has already been collected, however, it should be noted that the the data collection was prematurely stopped for the traces in this set. It is advised that when PeakTrace is used that the default run modules are modified to extended the data collection time by an additional two to three minutes. Such changes will further increase the performance of PeakTrace.

Figure 1. Quality score mapping of observed verses predicted quality scores for KB and PeakTrace basecalled traces. Observed quality values without observed errors or basecalls were plotted as the predicted quality value.
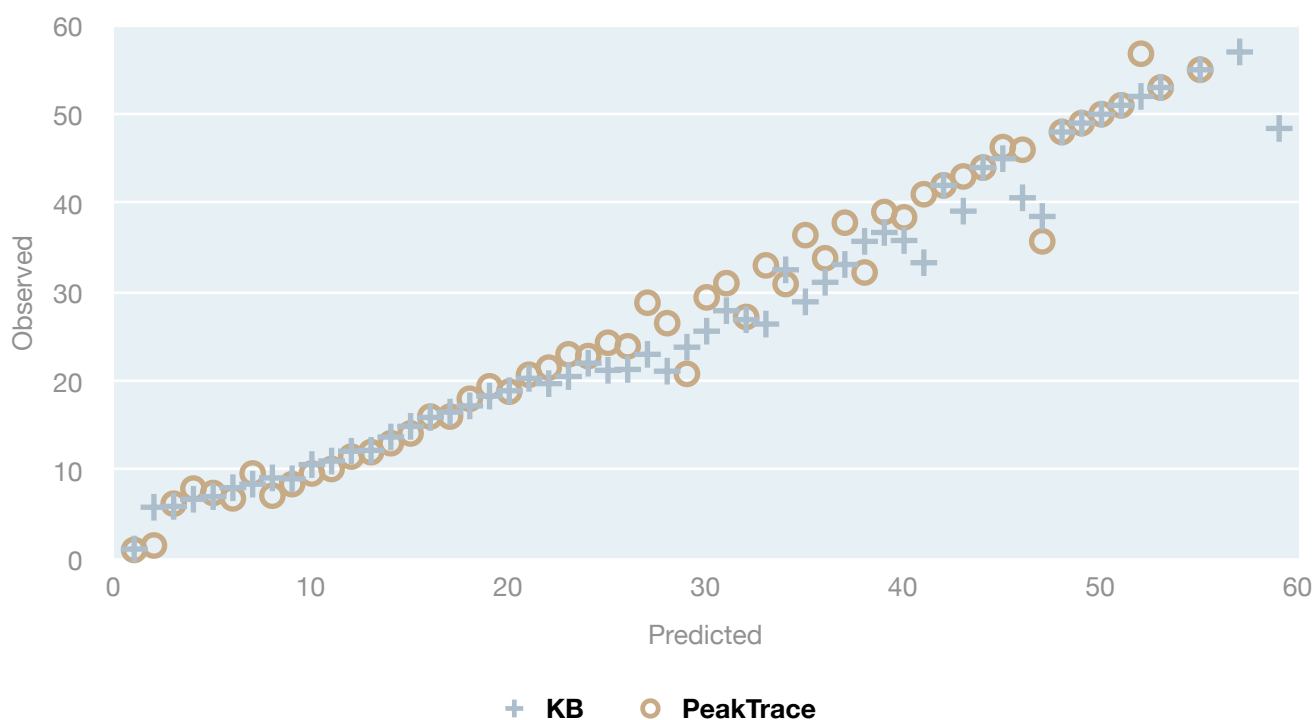
REFERENCES
1. Ewing B. & Green P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**(3):186-194.

AAGCTTGG C GT A NCATGG NCATAGCTGNTTCC TGNGTGAAATTG NATCCG
1160   1170   1180   1190   120
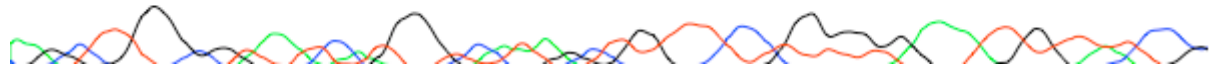
PEAKTRACE BASECALLER

VERSION	SKU	PRICE
PeakTrace:Box 500 (month)	PTB5	$150.00
PeakTrace:Box 1000 (month)	PTB10	$225.00
PeakTrace:Box 2000 (month)	PTB20	$400.00
PeakTrace:Box 5000 (month)	PTB50	$800.00
PeakTrace:Box 10000 (month)	PTB100	$1100.00
PeakTrace:Box 25000 (month)	PTB250	$2000.00
Auto PeakTrace RP 1000 Units	PTR10	$225.00
Auto PeakTrace RP 2000 Units	PTR20	$400.00
Auto PeakTrace RP 5000 Units	PTR50	$800.00
Auto PeakTrace RP 10000 Units	PTR100	$1100.00
Auto PeakTrace RP 25000 Units	PTR250	$2000.00

AGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCC
1160   1170   1180   1190   1200
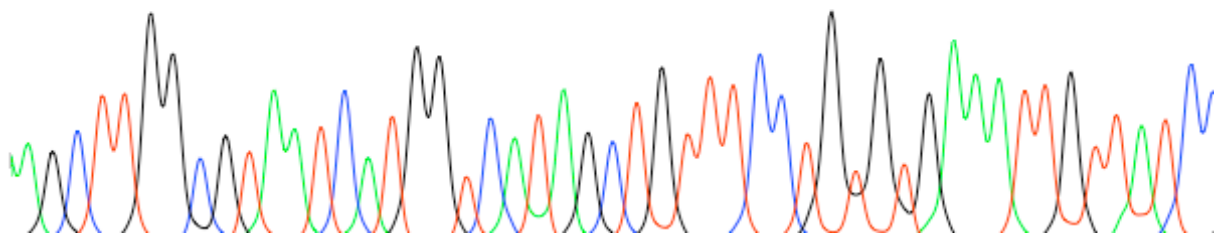
All price are in US$ and are subject to change without notice. Additional discount available. For Research Use Only. Not for use in diagnostic procedures.

Nucleics   T/+61390169373   12 Mill St Bendigo VIC 3550 Australia   www.nucleics.com