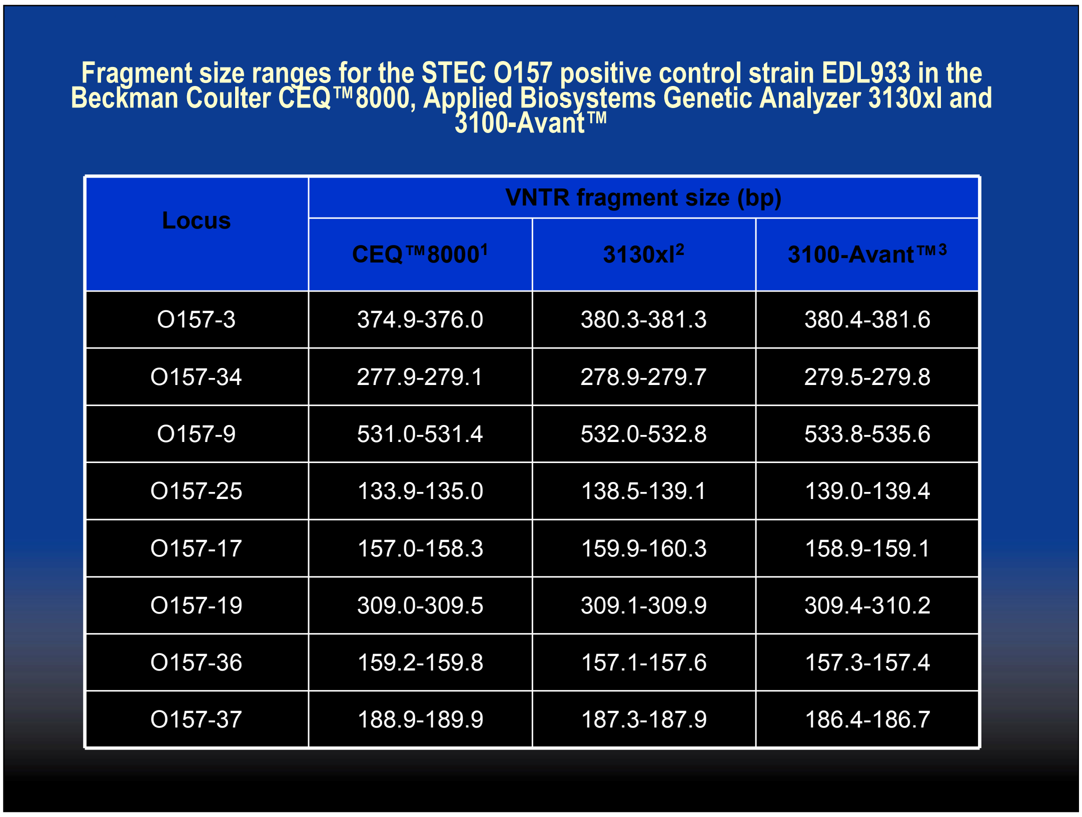
Analysis of MLVA data in BioNumerics

Eija Trees, Ph.D., D.V.M.

PulseNet Methods Development and Reference Unit

Enteric Diseases Laboratory Branch

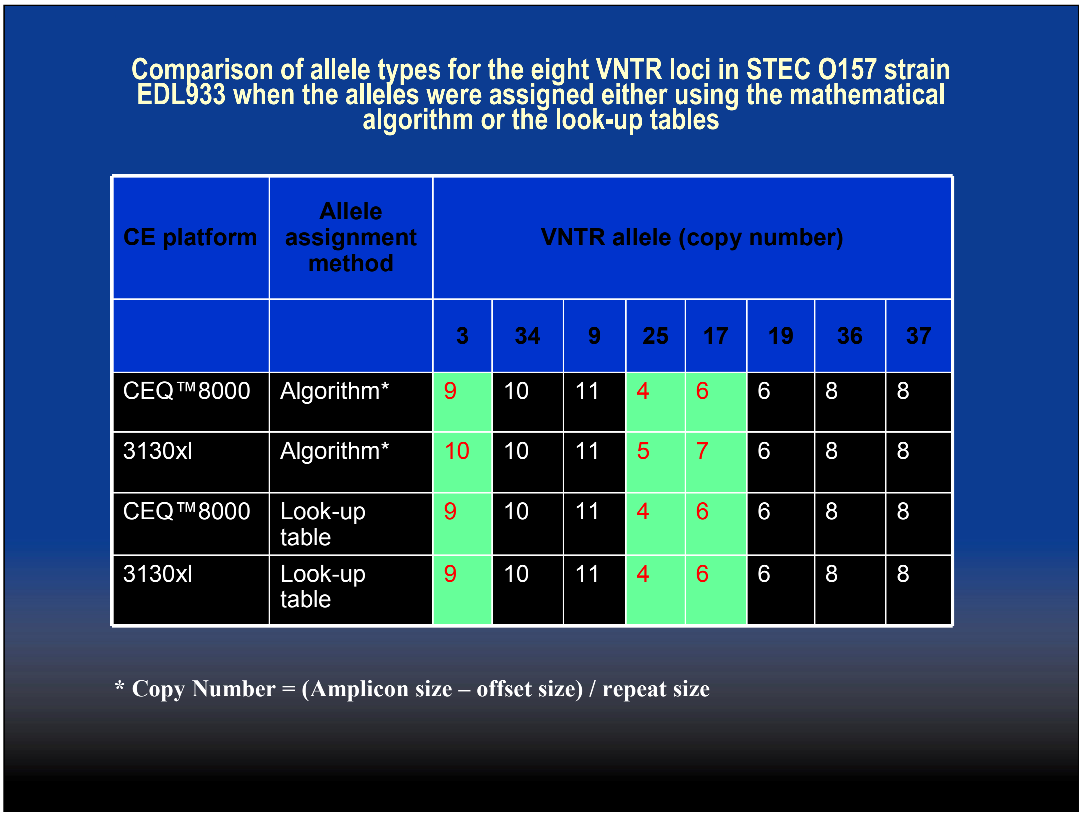
CDC, Atlanta, GA



Fragment size ranges for the STEC O157 positive control strain EDL933 in the Beckman Coulter CEQ™8000, Applied Biosystems Genetic Analyzer 3130xl and 3100-Avant™

Locus	VNTR fragment size (bp)		
	CEQ™8000[1]	3130xl[2]	3100-Avant™[3]
O157-3	374.9-376.0	380.3-381.3	380.4-381.6
O157-34	277.9-279.1	278.9-279.7	279.5-279.8
O157-9	531.0-531.4	532.0-532.8	533.8-535.6
O157-25	133.9-135.0	138.5-139.1	139.0-139.4
O157-17	157.0-158.3	159.9-160.3	158.9-159.1
O157-19	309.0-309.5	309.1-309.9	309.4-310.2
O157-36	159.2-159.8	157.1-157.6	157.3-157.4
O157-37	188.9-189.9	187.3-187.9	186.4-186.7

Sharing MLVA data between laboratories that have different capillary electrophoresis platforms is one of the biggest challenges since fragment sizing data generated with different platforms and even different versions of the same platform are not comparable with each other. The sizing discrepancies are caused by the different chemistries in terms of dyes, polymers and size standards employed by different platforms.



Comparison of allele types for the eight VNTR loci in STEC O157 strain EDL933 when the alleles were assigned either using the mathematical algorithm or the look-up tables

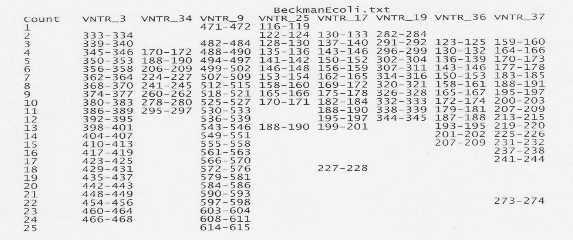
CE platform	**Allele assignment method**	**VNTR allele (copy number)**							
		3	**34**	**9**	**25**	**17**	**19**	**36**	**37**
CEQ™8000	Algorithm*	9	10	11	4	6	6	8	8
3130xl	Algorithm*	10	10	11	5	7	6	8	8
CEQ™8000	Look-up table	9	10	11	4	6	6	8	8
3130xl	Look-up table	9	10	11	4	6	6	8	8

* Copy Number = (Amplicon size – offset size) / repeat size

In some loci the sizing discrepancies between different platforms are large enough to result in a difference in the allele type if alleles are assigned based on the mathematical algorithm in which the offset (primer sequences and sequences between primers and the VNTR array) is deducted from the observed fragment size and the remaining number is divided by the repeat unit size. To address this issue, PulseNet USA has developed a BioNumerics analysis method that is based on platform specific look-up tables.

Look-up tables for the MLVA data analysis

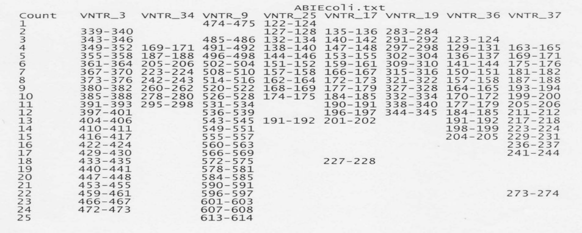
Beckman CEQ™8000



BeckmanEcoli.txt

Count	VNTR_3	VNTR_34	VNTR_9	VNTR_25	VNTR_17	VNTR_19	VNTR_36	VNTR_37
1			471-472	116-119				
2	333-334			122-124	130-133	282-284		
3	339-340		482-484	128-130	137-140	291-292	123-125	159-160
4	345-346	170-172	488-490	135-136	143-146	296-299	130-132	164-166
5	350-353	188-190	494-497	141-142	150-152	302-304	136-139	170-173
6	356-358	206-209	499-502	146-148	156-159	307-311	143-146	177-178
7	362-364	224-227	507-509	153-154	162-165	314-316	150-153	183-185
8	368-370	241-245	512-515	158-160	169-172	320-321	158-161	188-191
9	374-377	260-262	518-521	165-166	175-178	326-328	165-167	195-197
10	380-383	278-280	525-527	170-171	182-184	332-333	172-174	200-203
11	386-389	295-297	530-533		188-190	338-339	179-181	207-209
12	392-395		536-539		195-197	344-345	187-188	213-215
13	398-401		543-546	188-190	199-201		193-195	219-220
14	404-407		549-551				201-202	225-226
15	410-413		555-558				207-209	231-232
16	417-419		561-563					237-238
17	423-425		566-570					241-244
18	429-431		572-576		227-228			
19	435-437		579-581					
20	442-443		584-586					
21	448-449		590-593					
22	454-456		597-598					273-274
23	460-464		603-604					
24	466-468		608-611					
25			614-615					

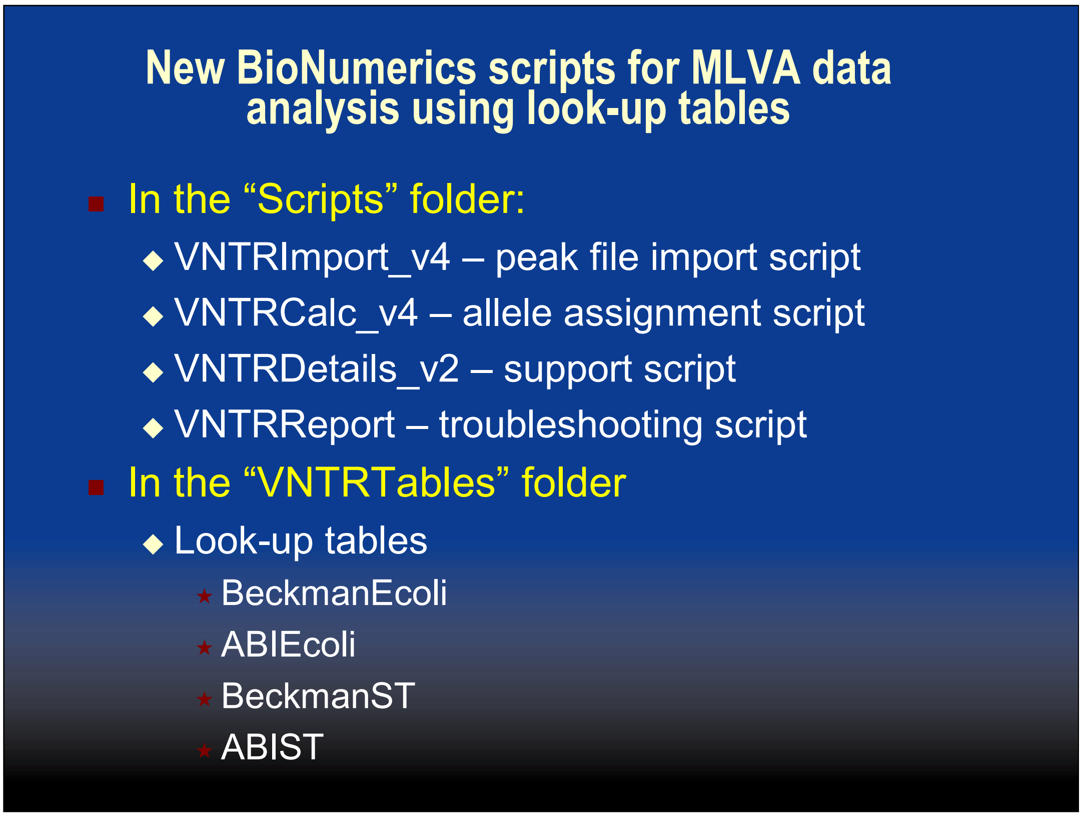
ABI Genetic Analyzer 3130xl



ABIEcoli.txt

Count	VNTR_3	VNTR_34	VNTR_9	VNTR_25	VNTR_17	VNTR_19	VNTR_36	VNTR_37
1			474-475	122-124				
2	339-340			127-128	135-136	283-284		
3	343-346		485-486	132-134	140-142	291-292	123-124	
4	349-352	169-171	491-492	138-140	147-148	297-298	129-131	163-165
5	355-358	187-188	496-498	144-146	153-155	302-304	136-137	169-171
6	361-364	205-206	502-504	151-152	159-161	309-310	141-144	175-176
7	367-370	223-224	508-510	157-158	166-167	315-316	150-151	181-182
8	373-376	242-243	514-516	162-164	172-173	321-322	157-158	187-188
9	380-382	260-262	520-522	168-169	177-179	327-328	164-165	193-194
10	385-388	278-280	526-528	174-175	184-185	332-334	170-172	199-200
11	391-393	295-298	531-534		190-191	338-340	177-179	205-206
12	397-401		536-539		196-197	344-345	184-185	211-212
13	404-406		543-545	191-192	201-202		191-192	217-218
14	410-411		549-551				198-199	223-224
15	416-417		555-557				204-205	229-231
16	422-424		560-563					236-237
17	429-430		566-569					241-244
18	433-435		572-575		227-228			
19	440-441		578-581					
20	447-448		584-585					
21	453-455		590-591					
22	459-461		596-597					273-274
23	466-467		601-603					
24	472-473		607-608					
25			613-614					

These tables are the allele assignment look-up tables for the *E. coli* O157 MLVA assay. The expected fragment size ranges for all alleles detected up to date for each eight VNTR loci are listed in the table. A BioNumerics script that refers to these tables for allele assignment was developed and has been implemented. When new alleles are detected, the tables can easily be modified by the database managers after the identity of the new allele has been confirmed by CDC PulseNet Central lab.

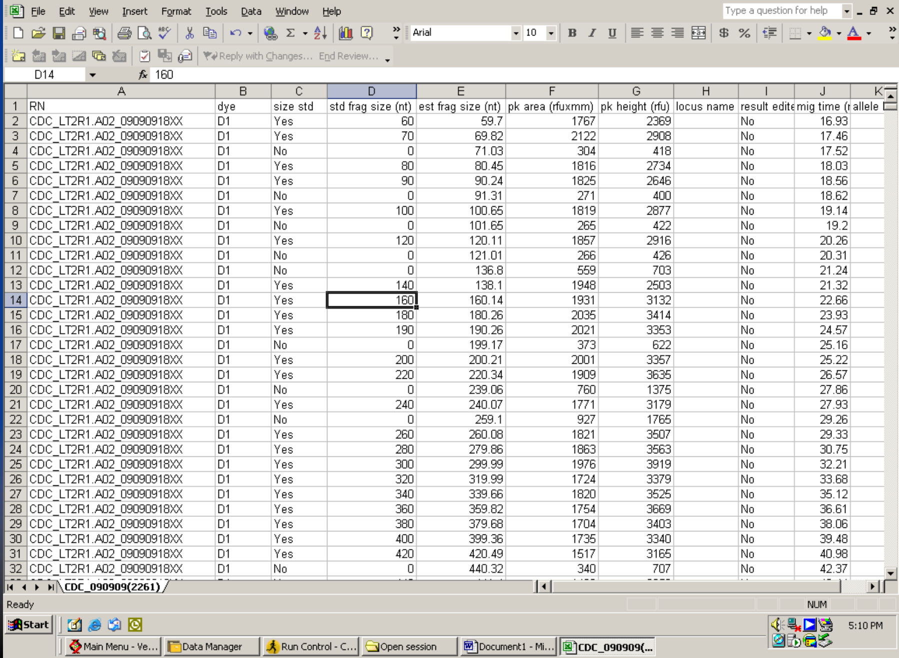


New BioNumerics scripts for MLVA data analysis using look-up tables

- In the “Scripts” folder:
    - VNTRImport_v4 – peak file import script
    - VNTRCalc_v4 – allele assignment script
    - VNTRDetails_v2 – support script
    - VNTRReport – troubleshooting script
- In the “VNTRTables” folder
    - Look-up tables
        - BeckmanEcoli
        - ABIEcoli
        - BeckmanST
        - ABIST

These are the MLVA scripts and look-up tables currently in use and available for PulseNet Participating labs. Eventually the scripts will be included in the PulseNet masterscripts. Until then the scripts can be requested by e-mailing the request to

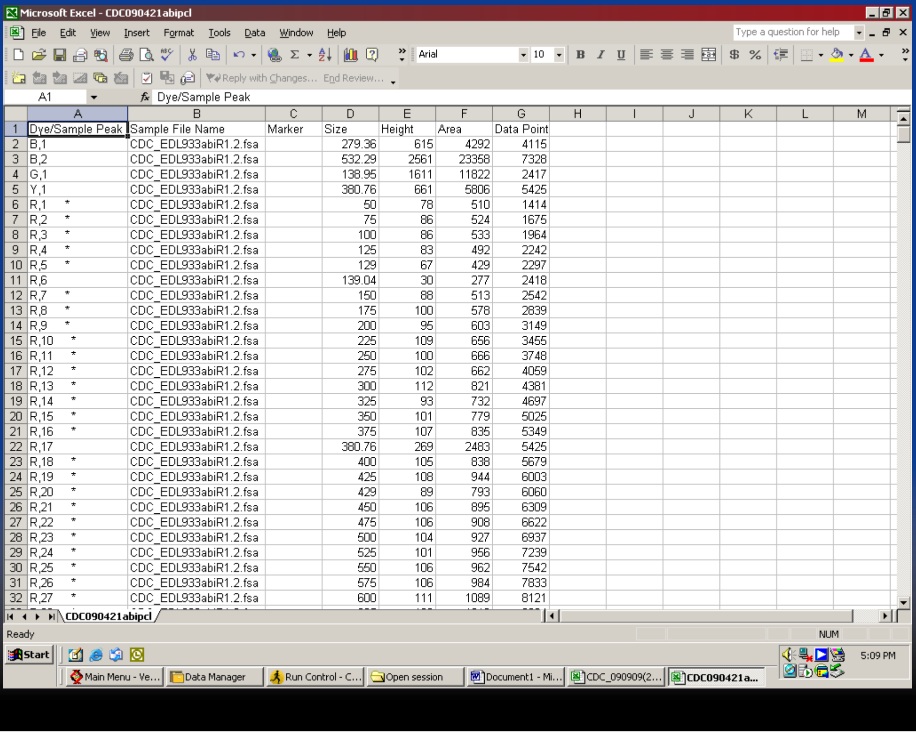
Required CEQ peak file format



	A	B	C	D	E	F	G	H	I	J	K
1	RN	dye	size std	std frag size (nt)	est frag size (nt)	pk area (rfuxmm)	pk height (rfu)	locus name	result edite	mig time (r	allele
2	CDC_LT2R1.A02_09090918XX	D1	Yes	60	59.7	1767	2369		No	16.93	
3	CDC_LT2R1.A02_09090918XX	D1	Yes	70	69.82	2122	2908		No	17.46	
4	CDC_LT2R1.A02_09090918XX	D1	No	0	71.03	304	418		No	17.52	
5	CDC_LT2R1.A02_09090918XX	D1	Yes	80	80.45	1816	2734		No	18.03	
6	CDC_LT2R1.A02_09090918XX	D1	Yes	90	90.24	1825	2646		No	18.56	
7	CDC_LT2R1.A02_09090918XX	D1	No	0	91.31	271	400		No	18.62	
8	CDC_LT2R1.A02_09090918XX	D1	Yes	100	100.65	1819	2877		No	19.14	
9	CDC_LT2R1.A02_09090918XX	D1	No	0	101.65	265	422		No	19.2	
10	CDC_LT2R1.A02_09090918XX	D1	Yes	120	120.11	1857	2916		No	20.26	
11	CDC_LT2R1.A02_09090918XX	D1	No	0	121.01	266	426		No	20.31	
12	CDC_LT2R1.A02_09090918XX	D1	No	0	136.8	559	703		No	21.24	
13	CDC_LT2R1.A02_09090918XX	D1	Yes	140	138.1	1948	2503		No	21.32	
14	CDC_LT2R1.A02_09090918XX	D1	Yes	160	160.14	1931	3132		No	22.66	
15	CDC_LT2R1.A02_09090918XX	D1	Yes	180	180.26	2035	3414		No	23.93	
16	CDC_LT2R1.A02_09090918XX	D1	Yes	190	190.26	2021	3353		No	24.57	
17	CDC_LT2R1.A02_09090918XX	D1	No	0	199.17	373	622		No	25.16	
18	CDC_LT2R1.A02_09090918XX	D1	Yes	200	200.21	2001	3357		No	25.22	
19	CDC_LT2R1.A02_09090918XX	D1	Yes	220	220.34	1909	3635		No	26.57	
20	CDC_LT2R1.A02_09090918XX	D1	No	0	239.06	760	1375		No	27.86	
21	CDC_LT2R1.A02_09090918XX	D1	Yes	240	240.07	1771	3179		No	27.93	
22	CDC_LT2R1.A02_09090918XX	D1	No	0	259.1	927	1765		No	29.26	
23	CDC_LT2R1.A02_09090918XX	D1	Yes	260	260.08	1821	3507		No	29.33	
24	CDC_LT2R1.A02_09090918XX	D1	Yes	280	279.86	1863	3563		No	30.75	
25	CDC_LT2R1.A02_09090918XX	D1	Yes	300	299.99	1976	3919		No	32.21	
26	CDC_LT2R1.A02_09090918XX	D1	Yes	320	319.99	1724	3379		No	33.68	
27	CDC_LT2R1.A02_09090918XX	D1	Yes	340	339.66	1820	3525		No	35.12	
28	CDC_LT2R1.A02_09090918XX	D1	Yes	360	359.82	1754	3669		No	36.61	
29	CDC_LT2R1.A02_09090918XX	D1	Yes	380	379.68	1704	3403		No	38.06	
30	CDC_LT2R1.A02_09090918XX	D1	Yes	400	399.36	1735	3340		No	39.48	
31	CDC_LT2R1.A02_09090918XX	D1	Yes	420	420.49	1517	3165		No	40.98	
32	CDC_LT2R1.A02_09090918XX	D1	No	0	440.32	340	707		No	42.37	

This is the required peak file format for the Beckman Coulter CEQ. All information fields shown on the slide need to be present for the import script to be able to read the file.

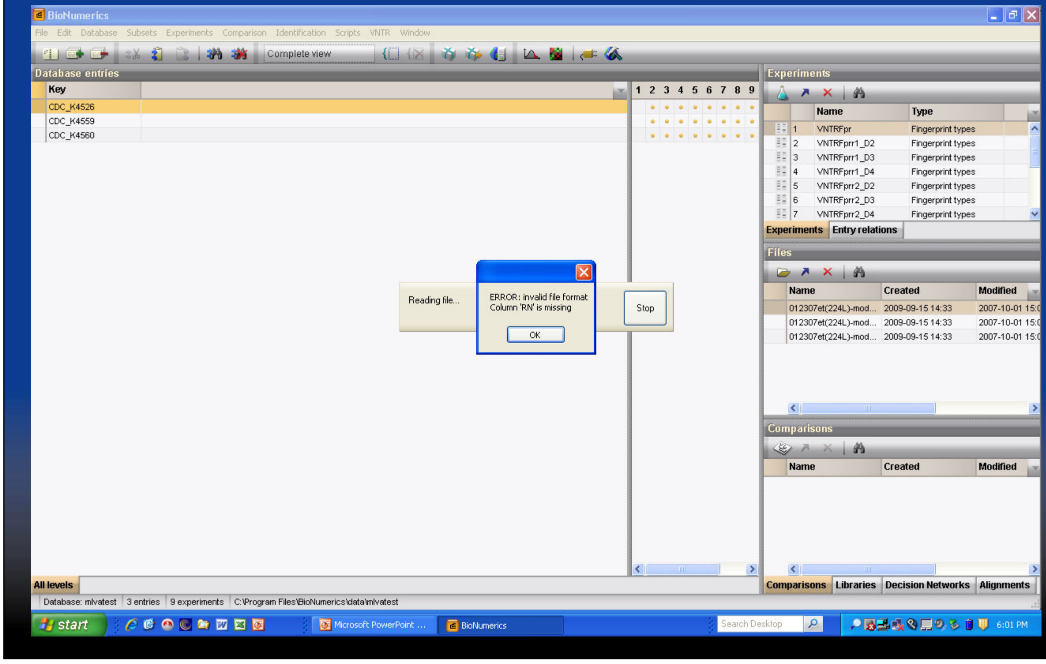
Required ABI peak file format



Dye/Sample Peak	Sample File Name	Marker	Size	Height	Area	Data Point
B,1	CDC_EDL933abiR1.2.fsa		279.36	615	4292	4115
B,2	CDC_EDL933abiR1.2.fsa		532.29	2561	23358	7328
G,1	CDC_EDL933abiR1.2.fsa		138.95	1611	11822	2417
Y,1	CDC_EDL933abiR1.2.fsa		380.76	661	5806	5425
R,1 *	CDC_EDL933abiR1.2.fsa		50	78	510	1414
R,2 *	CDC_EDL933abiR1.2.fsa		75	86	524	1675
R,3 *	CDC_EDL933abiR1.2.fsa		100	86	533	1964
R,4 *	CDC_EDL933abiR1.2.fsa		125	83	492	2242
R,5 *	CDC_EDL933abiR1.2.fsa		129	67	429	2297
R,6	CDC_EDL933abiR1.2.fsa		139.04	30	277	2418
R,7 *	CDC_EDL933abiR1.2.fsa		150	88	513	2542
R,8 *	CDC_EDL933abiR1.2.fsa		175	100	578	2839
R,9 *	CDC_EDL933abiR1.2.fsa		200	95	603	3149
R,10 *	CDC_EDL933abiR1.2.fsa		225	109	656	3455
R,11 *	CDC_EDL933abiR1.2.fsa		250	100	666	3748
R,12 *	CDC_EDL933abiR1.2.fsa		275	102	662	4059
R,13 *	CDC_EDL933abiR1.2.fsa		300	112	821	4381
R,14 *	CDC_EDL933abiR1.2.fsa		325	93	732	4697
R,15 *	CDC_EDL933abiR1.2.fsa		350	101	779	5025
R,16 *	CDC_EDL933abiR1.2.fsa		375	107	835	5349
R,17	CDC_EDL933abiR1.2.fsa		380.76	269	2483	5425
R,18 *	CDC_EDL933abiR1.2.fsa		400	105	838	5679
R,19 *	CDC_EDL933abiR1.2.fsa		425	108	944	6003
R,20 *	CDC_EDL933abiR1.2.fsa		429	89	793	6060
R,21 *	CDC_EDL933abiR1.2.fsa		450	106	895	6309
R,22 *	CDC_EDL933abiR1.2.fsa		475	106	908	6622
R,23 *	CDC_EDL933abiR1.2.fsa		500	104	927	6937
R,24 *	CDC_EDL933abiR1.2.fsa		525	101	956	7239
R,25 *	CDC_EDL933abiR1.2.fsa		550	106	962	7542
R,26 *	CDC_EDL933abiR1.2.fsa		575	106	984	7833
R,27 *	CDC_EDL933abiR1.2.fsa		600	111	1089	8121

This is the required peak file format for the Applied Biosystems Genetic Analyzer. All information fields shown on the slide need to be present for the import script to be able to read the file.

Failure to import a peak file into BioNumerics due to invalid file format



Key	1	2	3	4	5	6	7	8	9
CDC_K4526		●	●	●	●	●	●	●	●
CDC_K4559		●	●	●	●	●	●	●	●
CDC_K4560		●	●	●	●	●	●	●	●

	Name	Type
1	VNTRFpr	Fingerprint types
2	VNTRFprr1_D2	Fingerprint types
3	VNTRFprr1_D3	Fingerprint types
4	VNTRFprr1_D4	Fingerprint types
5	VNTRFprr2_D2	Fingerprint types
6	VNTRFprr2_D3	Fingerprint types
7	VNTRFprr2_D4	Fingerprint types

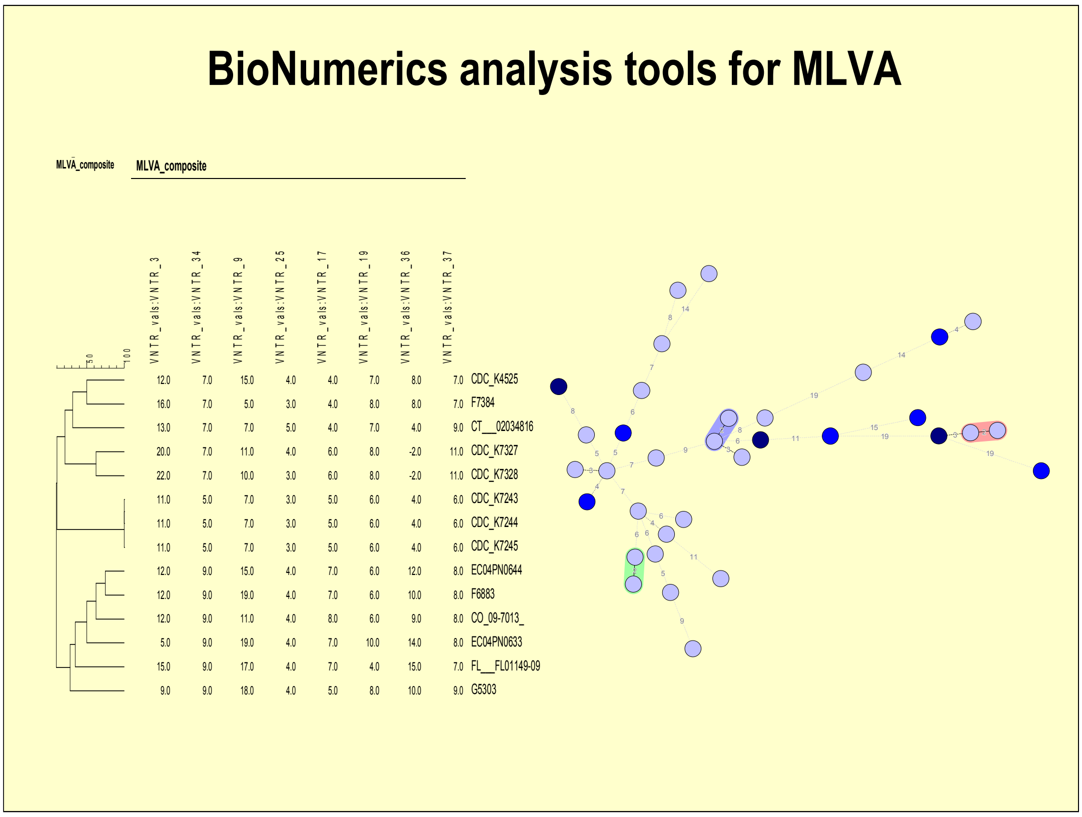
Name	Created	Modified
012307et(224L)-mod...	2009-09-15 14:33	2007-10-01 15:0
012307et(224L)-mod...	2009-09-15 14:33	2007-10-01 15:0
012307et(224L)-mod...	2009-09-15 14:33	2007-10-01 15:0

Name	Created	Modified

```
ERROR: invalid file format
Column 'RN' is missing
```

Database: mlvatest | 3 entries | 9 experiments | C:\Program Files\BioNumerics\data\mlvatest

If the peak file is in wrong format, the above error message will be displayed by BioNumerics.



BioNumerics analysis tools for MLVA

MLVA_composite

MLVA_composite

VNTR_vals:VNTR_3	VNTR_vals:VNTR_34	VNTR_vals:VNTR_9	VNTR_vals:VNTR_25	VNTR_vals:VNTR_17	VNTR_vals:VNTR_19	VNTR_vals:VNTR_36	VNTR_vals:VNTR_37	
12.0	7.0	15.0	4.0	4.0	7.0	8.0	7.0	CDC_K4525
16.0	7.0	5.0	3.0	4.0	8.0	8.0	7.0	F7384
13.0	7.0	7.0	5.0	4.0	7.0	4.0	9.0	CT___02034816
20.0	7.0	11.0	4.0	6.0	8.0	-2.0	11.0	CDC_K7327
22.0	7.0	10.0	3.0	6.0	8.0	-2.0	11.0	CDC_K7328
11.0	5.0	7.0	3.0	5.0	6.0	4.0	6.0	CDC_K7243
11.0	5.0	7.0	3.0	5.0	6.0	4.0	6.0	CDC_K7244
11.0	5.0	7.0	3.0	5.0	6.0	4.0	6.0	CDC_K7245
12.0	9.0	15.0	4.0	7.0	6.0	12.0	8.0	EC04PN0644
12.0	9.0	19.0	4.0	7.0	6.0	10.0	8.0	F6883
12.0	9.0	11.0	4.0	8.0	6.0	9.0	8.0	CO_09-7013_
5.0	9.0	19.0	4.0	7.0	10.0	14.0	8.0	EC04PN0633
15.0	9.0	17.0	4.0	7.0	4.0	15.0	7.0	FL___FL01149-09
9.0	9.0	18.0	4.0	5.0	8.0	10.0	9.0	G5303

In BioNumerics, MLVA data can be analyzed with all tools available for character data. Data can be visualized using dendrograms and minimum spanning trees.