INSTITUTE of APPLIED GENETICS
Department of Molecular and Medical Genetics
UNTHSC Research and Development Laboratory

STRait Razor Analysis Minimization Protocol

Purpose

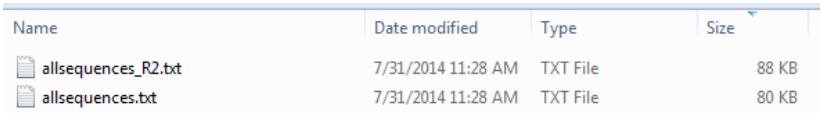
In order to reduce the amount of processing time required for STRait Razor Analysis (SRA) to complete various tasks, this short note was written detailing the cells to be modified. Those less comfortable with Microsoft Excel may contact Jonathan.King@unthsc.edu in order to alter the workbook to suit individual requirements.

*I've broken down this note into sections for **Sorting** and **Exporting**, respectively. It's important first to understand why the workbook takes longer (~5-10 minutes) to complete seemingly small tasks. The workbook is designed to process large-scale multiplexes capable of producing tremendous amounts of data. In order to account for the possible large amounts of data, both the **Standard** and **Lite** versions are equipped with formulas above and beyond what is likely to be needed in non-mixture or even two-three contributor mixture samples. These "extra" formulas may be removed easily using this technical note.*

Special thanks to users at ESR for their feedback essential to the preparation of this document.

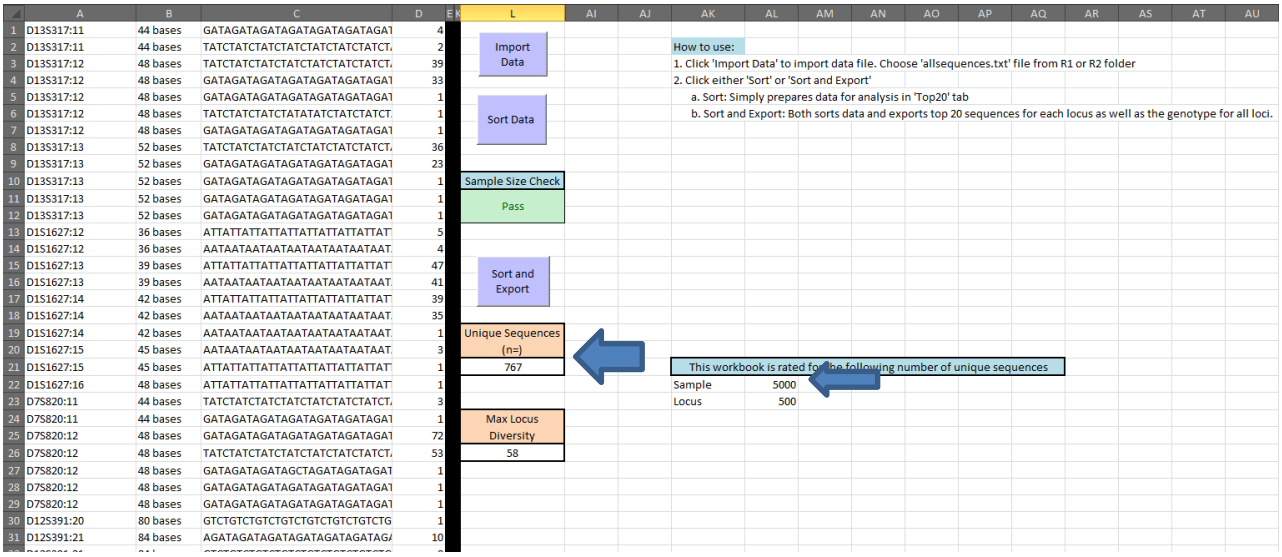
Sort

1. Determine the level of unique samples in your dataset by opening the largest 3-5 'allsequences.txt' files in your dataset by sorting files by size.



Name	Date modified	Type	Size
allsequences_R2.txt	7/31/2014 11:28 AM	TXT File	88 KB
allsequences.txt	7/31/2014 11:28 AM	TXT File	80 KB

2. These example text files contain 868 and 767 rows of text, respectively. These numbers can be noted upon import as in the figure below.



Note: The Lite edition of SRA is capable of analyzing up to 5000 rows of unique sequences (as seen above); however, this may be a larger workbook than your dataset requires. If so, the following steps can be done to decrease processing time. As a safety measure, the sample size check will alert the user to samples that are too large for the current workbook.

3. If, for example, five samples have 1250, 1200, 1000, 900, and 750 unique sequences each, the pertinent thing to do would be to prepare for more diverse samples while not wasting time. By taking the average plus three standard deviations, we find that ~1642 should account for most samples.

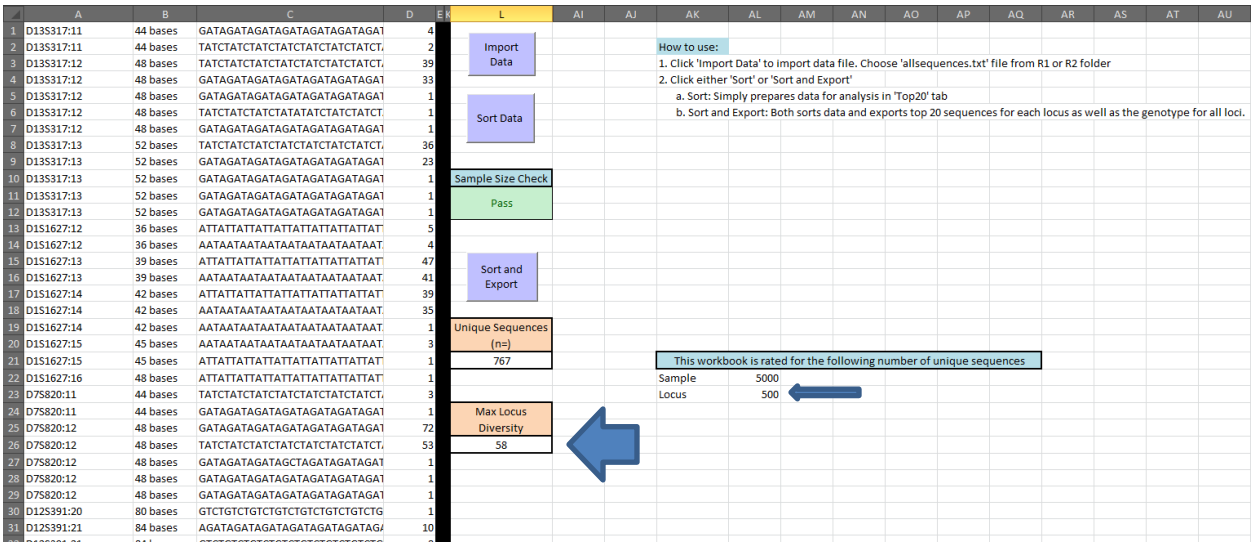
4. Highlight unnecessary rows (e.g., 1651:50000) and delete these rows entirely.

   Note: It's important to delete the rows and not simply clear the cells. Hidden cells, which affect performance, will remain otherwise.

5. This should greatly reduce the time needed to sort and prepare data for additional analyses.

Export

1. To increase export efficiency, first determine max locus diversity from a number of samples.

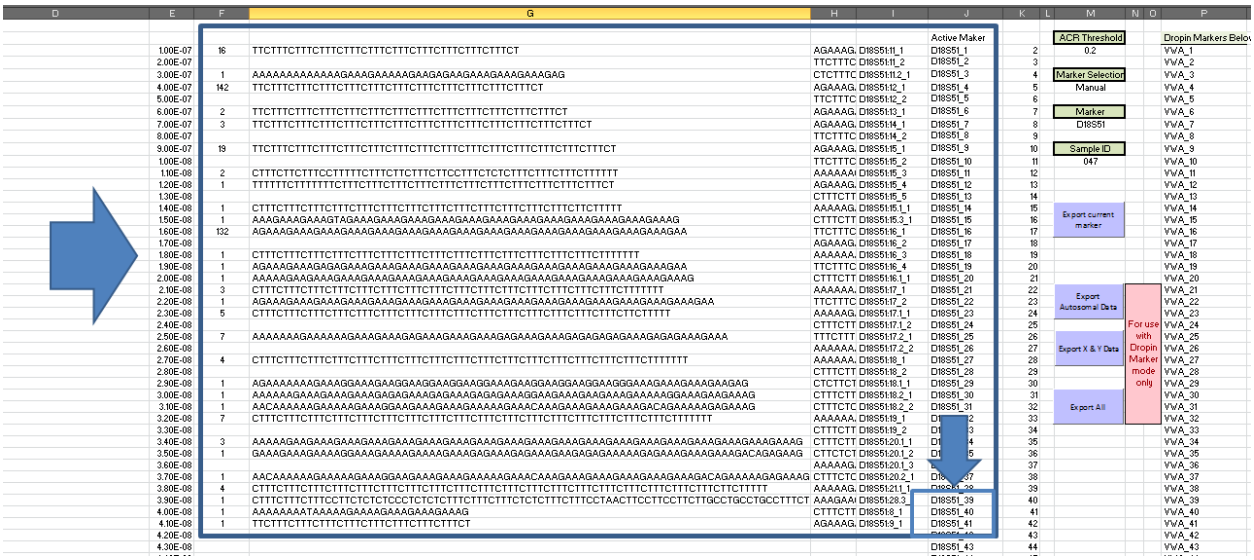


Note: The example workbook, Lite edition, allows for 500 unique sequences per locus. While this is able to capture complex mixtures, it is beyond the scope of most users data.

2. Once the level of locus diversity is determined (e.g., 120), select the 'Top20' worksheet or tab.

3. Unhide columns A:K

4. The level of locus diversity is found in column J for this locus 41 unique sequences were found.



D	E	F	G	H	I	J	K	L	M	N	O	P			
						Active Maker			ACR Threshold			Dropin Markers Belo			
	1.00E-07	16	TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT	AGAAAG,	D18S51:11_1	D18S51_1	2		0.2			VWA_1			
	2.00E-07			TTCTTTC	D18S51:11_2	D18S51_2	3					VWA_2			
	3.00E-07	1	AAAAAAAAAAAAAGAAAGAAAAAGAAGAGAAGAAAGAAAGAAAGAAAAG	CTCTTTC	D18S51:11.2_1	D18S51_3	4		Marker Selection			VWA_3			
	4.00E-07	142	TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT	AGAAAG,	D18S51:12_1	D18S51_4	5		Manual			VWA_4			
	5.00E-07			TTCTTTC	D18S51:12_2	D18S51_5	6					VWA_5			
	6.00E-07	2	TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT	AGAAAG,	D18S51:13_1	D18S51_6	7		Marker			VWA_6			
	7.00E-07	3	TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT	AGAAAG,	D18S51:14_1	D18S51_7	8		D18S51			VWA_7			
	8.00E-07			TTCTTTC	D18S51:14_2	D18S51_8	9					VWA_8			
	9.00E-07	19	TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT	AGAAAG,	D18S51:15_1	D18S51_9	10		Sample ID			VWA_9			
	1.00E-08			TTCTTTC	D18S51:15_2	D18S51_10	11		047			VWA_10			
	1.10E-08	2	CTTTCTTCTTTCCTTTTTCTTTCTTCTTTCTTCCTTTCTCTCTTTCTTTCTTTTTTT	AAAAAAA	D18S51:15_3	D18S51_11	12					VWA_11			
	1.20E-08	1	TTTTTTCTTTTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT	AGAAAG,	D18S51:15_4	D18S51_12	13					VWA_12			
	1.30E-08			CTTTCTT	D18S51:15_5	D18S51_13	14					VWA_13			
	1.40E-08	1	CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTTT	AAAAAG,	D18S51:15.1_1	D18S51_14	15		Export current marker			VWA_14			
	1.50E-08	1	AAAGAAAGAAAGTAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	CTTTCTT	D18S51:15.3_1	D18S51_15	16					VWA_15			
	1.60E-08	132	AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	TTCTTTC	D18S51:16_1	D18S51_16	17					VWA_16			
	1.70E-08			AGAAAG,	D18S51:16_2	D18S51_17	18					VWA_17			
	1.80E-08	1	CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTTTTT	AAAAAA,	D18S51:16_3	D18S51_18	19					VWA_18			
	1.90E-08	1	AGAAAGAAAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	TTCTTTC	D18S51:16_4	D18S51_19	20					VWA_19			
	2.00E-08	1	AAAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	CTTTCTT	D18S51:16.1_1	D18S51_20	21					VWA_20			
	2.10E-08	3	CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTTTTT	AAAAAA,	D18S51:17_1	D18S51_21	22		Export Autosomal Data			VWA_21			
	2.20E-08	1	AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	TTCTTTC	D18S51:17_2	D18S51_22	23					VWA_22			
	2.30E-08	5	CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTTTTT	AAAAAG,	D18S51:17.1_1	D18S51_23	24					VWA_23			
	2.40E-08			CTTTCTT	D18S51:17.1_2	D18S51_24	25				For use with Dropin Marker mode only	VWA_24			
	2.50E-08	7	AAAAAAAGAAAAAAGAAAGAAAGAGAAAGAAAGAAAGAAAGAGAAAGAAAGAGAGAGAAAGAGAAAGAAA	TTTCTTT	D18S51:17.2_1	D18S51_25	26					VWA_25			
	2.60E-08			AAAAAA,	D18S51:17.2_2	D18S51_26	27		Export X & Y Data			VWA_26			
	2.70E-08	4	CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTTTTT	AAAAAA,	D18S51:18_1	D18S51_27	28					VWA_27			
	2.80E-08			CTTTCTT	D18S51:18_2	D18S51_28	29					VWA_28			
	2.90E-08	1	AGAAAAAAAGAAAGGAAAGAAGGAAGGAAGGAAGGAAAGAAGGAAGGAAGGAAAGAAAGAAAGAAGAG	CTCTTCT	D18S51:18.1_1	D18S51_29	30					VWA_29			
	3.00E-08	1	AAAAAAGAAAGAAAGAAAGAGAAGAAAGAGAAAGAGAGAAAGGAAGAAAGAAGAAAAGAAAGAAAAAGGAAAGAAGAAAG	CTTTCTT	D18S51:18.2_1	D18S51_30	31		Export All			VWA_30			
	3.10E-08	1	AACAAAAAAGAAAAAGAAAGGAAGAAAGAAAGAAAAGAAACAAAGAAAGAAAGAAAGACAGAAAAAGAGAAAG	CTTTCTC	D18S51:18.2_2	D18S51_31	32					VWA_31			
	3.20E-08	7	CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTTTTT	AAAAAA,	D18S51:19_1	D1 2	33					VWA_32			
	3.30E-08			CTTTCTT	D18S51:19_2	D1 3	34					VWA_33			
	3.40E-08	3	AAAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	CTTTCTT	D18S51:20.1_1	D1 4	35					VWA_34			
	3.50E-08	1	GAAAGAAAGAAAAGGAAAGAAAAGAAAAGAAAGAAAGAGAAAGAGAAAAGAAGAGAGAAAAAGAGAAAGAAAGAAAGACAGAGAAG	CTTCTCT	D18S51:20.1_2	D1 5	36					VWA_35			
	3.60E-08			AAAAAG,	D18S51:20.1_3		37					VWA_36			
	3.70E-08	1	AACAAAAAAGAAAAAGAAAGGAAGAAAGAAAGAAAAAGAAACAAAGAAAGAAAGAAAGAAAGAAAGACAGAAAAAGAGAAAG	CTTTCTC	D18S51:20.2_1	D1 7	38					VWA_37			
	3.80E-08	4	CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTT	AAAAAG,	D18S51:21.1_1	D1 8	39					VWA_38			
	3.90E-08	1	CTTTCTTTCTTTCCTTCTCTCTCCCTCTCTCTTTCTTTCTTTCTTTCTCTCTTTCCTAACTTCCTTCCTTCTTGCCTGCCTGCCTTTCT	AAAGAAA	D18S51:28.3_1	D18S51_39	40					VWA_39			
	4.00E-08	1	AAAAAAAATAAAAAGAAAAGAAAAGAAAGAAAG	CTTTCTT	D18S51:8_1	D18S51_40	41					VWA_40			
	4.10E-08	1	TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT	AGAAAG,	D18S51:9_1	D18S51_41	42					VWA_41			
	4.20E-08						43					VWA_42			
	4.30E-08					D18S51_43	44					VWA_43			

5. In this example, 120 sequences were found to be sufficient for each locus assayed. To decrease exporting time, select cells E123:J500 and clear the cell contents.

   Note: Do NOT delete the cells. Clearing the content from these cells is sufficient.