

Accurate genome wide multiple alignments and reconstruction in mammal

Downloads of the programs and the entire dataset are available at <http://www.ebi.ac.uk/~jherrero/docs/epo/>

Ancestral repeat analysis

List of Ancestral repeats

We choosed the ancestral repeats using the following criteria:

- they were Type II transposons
- they were present between 300 and 50000 times in every genome
- the number of repeats for each species fell within the ratio:
 - mouse/human between 0.2 and 0.7
 - rat/human between 0.2 and 0.7
 - dog/human between 0.2 and 1.0
 - cow/human between 0.2 and 1.0

Supplementary Table 1: List of ancestral repeats. A table showing the list of Ancestral Repeats used in the assessment of the alignments as well as the number of times we find each repeat in each genome

Ancestral repeat	Presence in each genome				
	Human	Mouse	Rat	Dog	Cow
Charlie1	6148	3494	2618	5171	5815
Charlie4	1630	775	557	1239	1543
Cheshire	880	456	393	654	670
MER5A	40768	12528	9027	30932	34870
MER20	18777	6358	4484	14297	16808
MER45B	1314	699	446	1050	1130
MER58B	7864	2879	2034	6251	7209
MER82	3634	1917	1342	2853	3176
MER119	1333	441	321	1060	1165

Ancestral repeat correspondence

We found that some of these repeats were aligned to homologous ancestral repeats in the other species. In order to improve the coverage of our analysis, we compared and aligned the consensus sequence of the ancestral repeats in order to build the following correspondence table.

Some of these cases correspond to perfect matches between these regions while in other cases we used exonerate (Slater and Birney, BMC Bioinformatics 2005, 6:31) to align and match the positions from one ancestral repeat to the other.

Supplementary Table 2: Correspondence between ancestral repeats.

Correspondence between ancestral repeats			
Ancestral repeat	Region	Aligned ancestral repeat	Region
Charlie1	1 – 115	Charlie1a	1 – 115
	1427 – 2761		121 – 1455
	1 – 171	Charlie1b	1 – 171
	2416 – 2761		178 – 523
Charlie4	1 – 289	Charlie4a	1 – 289
	1745 – 1961		292 – 508
Cheshire	1 – 2420	MER58	1 – 2420
MER5A	36 – 189	MER58A	25 – 178
MER45B	1 – 1040	MER45	1 – 1040
MER58B	1 – 341	MER58	1 – 341
	1 – 46	MER58A	1 – 46
	175 – 341		58 – 224

Results by ancestral repeat

The following tables contain the details of the ancestral repeat assessment for each ancestral repeat

Supplementary Table 3: Ancestral repeat assessment for Charlie1 repeats

Charlie1						
Alignment	Results in bp			Results as %		
	Full matches	Partial matches	Mismatches	Full matches	Partial matches	Mismatches
Mercator+Mlagan	194522	124936	282737	32.3	20.75	46.95
Mercator+Pecan	318892	145283	209119	47.36	21.58	31.06
MutliZ	330213	283135	239281	38.73	33.21	28.06
Enredo+Pecan	463561	174291	266404	51.26	19.27	29.46

Supplementary Table 4: Ancestral repeat assessment for Charlie4 repeats

Charlie4						
Alignment	Results in bp			Results as %		
	Full matches	Partial matches	Mismatches	Full matches	Partial matches	Mismatches
Mercator+Mlagan	22003	23202	35384	27.3	28.79	43.91
Mercator+Pecan	33178	28465	26220	37.76	32.4	29.84
MutliZ	40109	50692	30452	33.08	41.81	25.11
Enredo+Pecan	49550	40071	33656	40.19	32.5	27.3

Supplementary Table 5: Ancestral repeat assessment for Cheshire repeats

Cheshire						
Alignment	Results in bp			Results as %		
	Full matches	Partial matches	Mismatches	Full matches	Partial matches	Mismatches
Mercator+Mlagan	60198	26616	67261	39.07	17.27	43.65
Mercator+Pecan	91762	29785	47492	54.28	17.62	28.1
MutliZ	83158	62799	50070	42.42	32.04	25.54
Enredo+Pecan	132022	31994	57126	59.7	14.47	25.83

Supplementary Table 6: Ancestral repeat assessment for MER5A repeats

MER5A						
Alignment	Results in bp			Results as %		
	Full matches	Partial matches	Mismatches	Full matches	Partial matches	Mismatches
Mercator+Mlagan	465640	585880	719547	26.29	33.08	40.63
Mercator+Pecan	684916	649317	540847	36.53	34.63	28.84
MutliZ	960529	1296187	913138	30.3	40.89	28.81
Enredo+Pecan	1127082	934530	800396	39.38	32.65	27.97

Supplementary Table 7: Ancestral repeat assessment for MER20 repeats

MER20						
Alignment	Results in bp			Results as %		
	Full matches	Partial matches	Mismatches	Full matches	Partial matches	Mismatches
Mercator+Mlagan	401733	385751	537773	30.31	29.11	40.58
Mercator+Pecan	603321	440762	363789	42.85	31.31	25.84
MutliZ	779070	920097	573600	34.28	40.48	25.24
Enredo+Pecan	946760	607954	513671	45.77	29.39	24.83

Supplementary Table 8: Ancestral repeat assessment for MER45B repeats

MER45B						
Alignment	Results in bp			Results as %		
	Full matches	Partial matches	Mismatches	Full matches	Partial matches	Mismatches
Mercator+Mlagan	50040	26042	62612	36.08	18.78	45.14
Mercator+Pecan	78959	31411	43186	51.42	20.46	28.12
MutliZ	126009	116976	89031	37.95	35.23	26.82
Enredo+Pecan	169425	58071	77956	55.47	19.01	25.52

Supplementary Table 9: Ancestral repeat assessment for MER58B repeats

MER58B						
Alignment	Results in bp			Results as %		
	Full matches	Partial matches	Mismatches	Full matches	Partial matches	Mismatches
Mercator+Mlagan	205013	182741	275037	30.93	27.57	41.5
Mercator+Pecan	311873	206620	189539	44.05	29.18	26.77
MutliZ	388237	445354	262720	35.41	40.62	23.96
Enredo+Pecan	475853	279305	258506	46.94	27.55	25.5

Supplementary Table 10: Ancestral repeat assessment for MER82 repeats

MER82						
Alignment	Results in bp			Results as %		
	Full matches	Partial matches	Mismatches	Full matches	Partial matches	Mismatches
Mercator+Mlagan	173645	92531	188157	38.22	20.37	41.41
Mercator+Pecan	276265	99976	114748	56.27	20.36	23.37
MutliZ	345461	281915	184235	42.56	34.74	22.7
Enredo+Pecan	451728	137978	159015	60.33	18.43	21.24

Supplementary Table 11: Ancestral repeat assessment for MER119 repeats

MER119						
Alignment	Results in bp			Results as %		
	Full matches	Partial matches	Mismatches	Full matches	Partial matches	Mismatches
Mercator+Mlagan	42513	38200	58790	30.47	27.38	42.14
Mercator+Pecan	64767	44812	40368	43.19	29.89	26.92
MutliZ	83597	94099	61485	34.95	39.34	25.71
Enredo+Pecan	105718	62119	57879	46.84	27.52	25.64