

# Supplementary material for ‘Cactus: algorithms for genome multiple sequence alignment’

Benedict Paten<sup>\*1</sup>, Dent Earl<sup>1</sup>, Ngan Nguyen<sup>1</sup>, Mark Diekhans<sup>1</sup>, Daniel Zerbino<sup>1</sup>, David Haussler<sup>\*\*1</sup>

May 27, 2011

Corresponding authors \* benedict@soe.ucsc.edu \*\* haussler@soe.ucsc.edu

(1) Center for Biomolecular Science and Engineering, University of California Santa Cruz, CA,  
USA

Figures 1 and 2 show a macro view of the evolution of the Evolver mammals and primates simulations, respectively. Figure 3 shows the effect of disabling the repetitive transposition operation on the Evolver mammals simulation. Figures 4-8 show different examples of non-conserved primate exon alignments. Tables 1-8 show pairwise comparisons for the individual species in the Evolver mammals and Evolver primates simulations. Tables 9 and 10 detail exons which are conserved by Cactus but not Multiz, and vice versa.

## References

- [Fujita et al., 2011] Fujita, P. A., Rhead, B., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Cline, M. S., Goldman, M., Barber, G. P., Clawson, H., Coelho, A., *et al.*, 2011. The ucsc genome browser database: update 2011. *Nucleic Acids Res*, **39**(Database issue):D876–82.

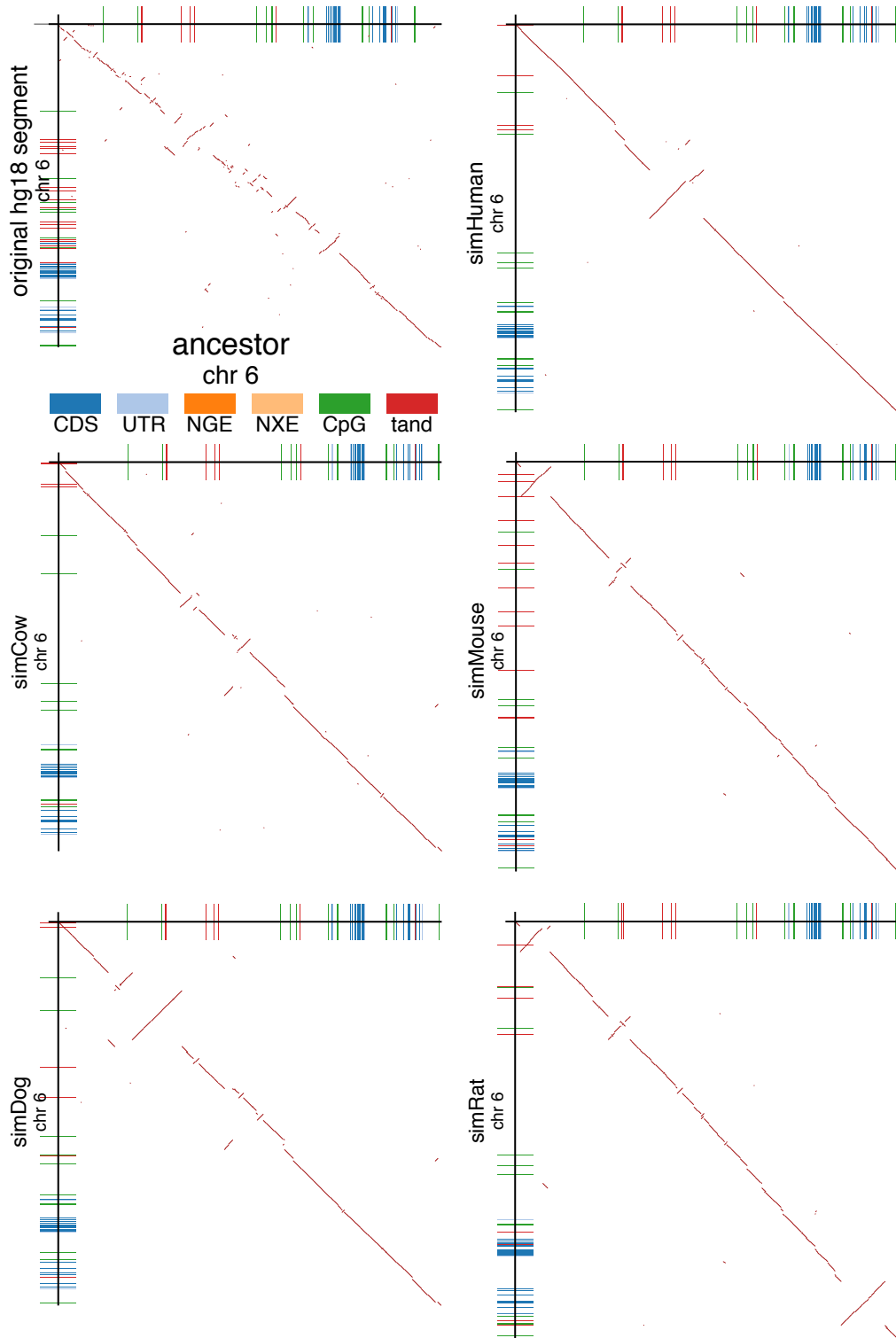


Figure 1: Dot-plots showing the large-scale events in the Evolver mammals simulation. In every plot the x axis is the MRCA. The y axes clockwise from top left: the original root genome, the simulated human, the simulated mouse, the simulated rat, the simulated dog and the simulated cow. Annotations are shown along the axes according to the colour key. CDS: coding sequence. UTR: untranslated region. NGE: non-genic constrained element, NXE = non-exonic (e.g. intronic ) constrained element, CpG: CpG island, tand: Tandem duplication.

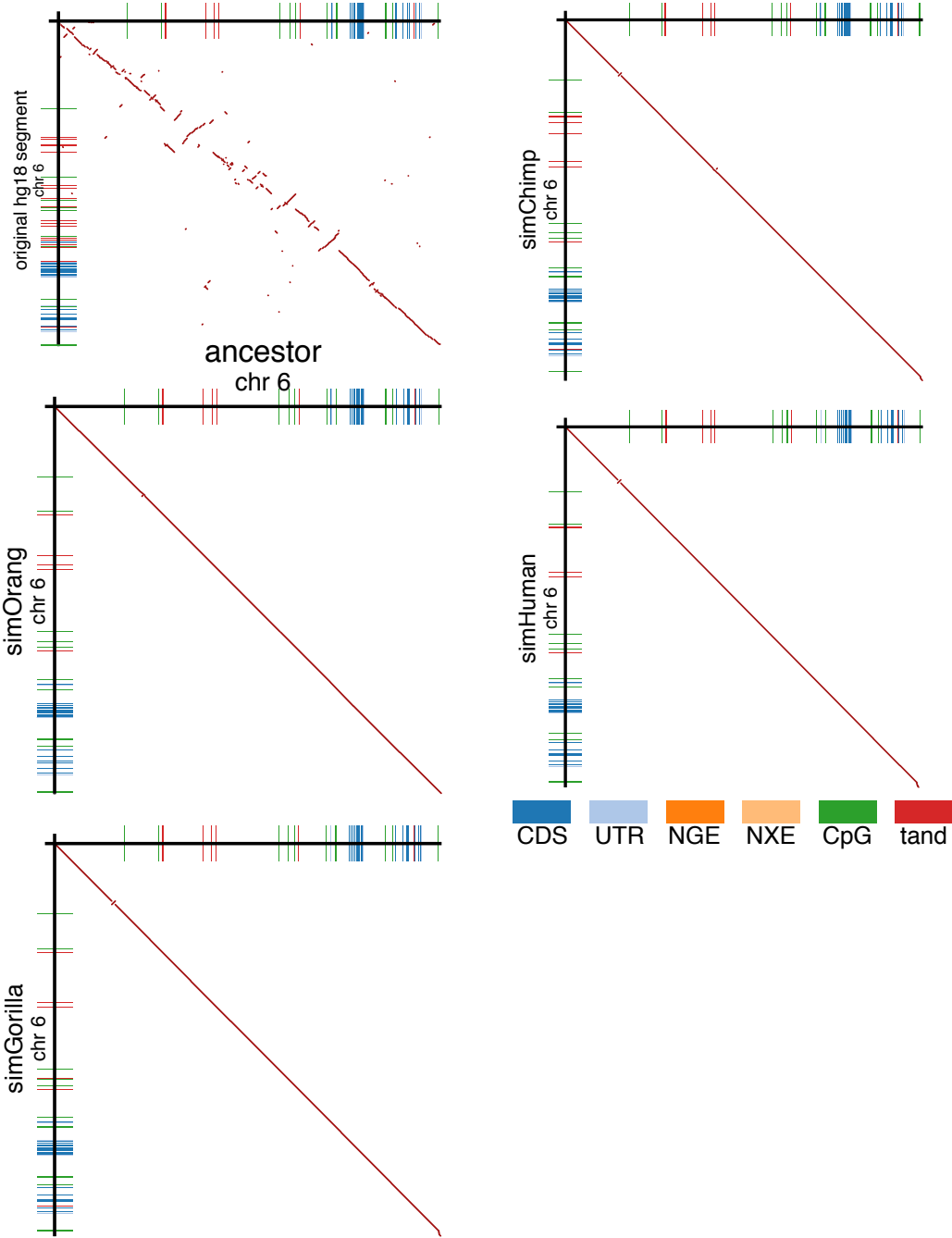


Figure 2: Dot-plots showing the large-scale events in the Evolver primates simulation. In every plot the x axis is the MRCA. The y axes clockwise from top left: the original root genome, the simulated chimp, the simulated human, the simulated gorilla and the simulated orangutan. Annotations are shown along the axes according to the colour key, same as in Figure 1.

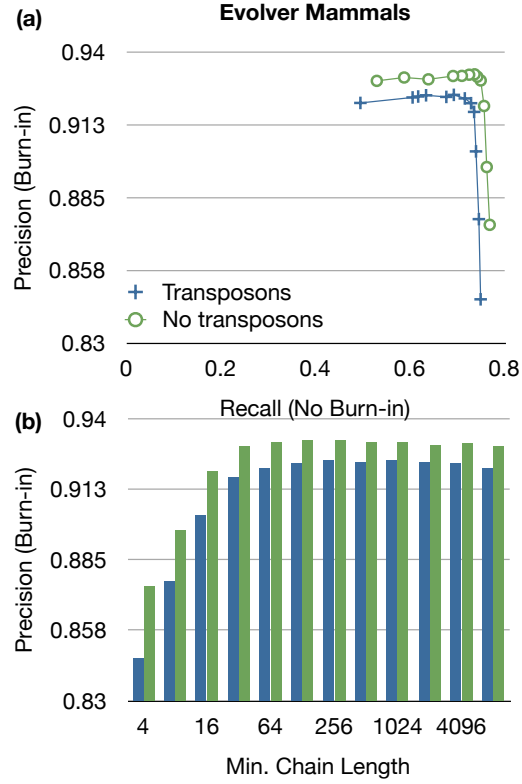


Figure 3: Comparing the effect of removing the repetitive transposition operation on two independent runs of the Evolver mammals simulation, one including and one without the operation. (a) A precision-recall plot of the predicted Cactus alignments showing the effect of altering  $\alpha$  in powers of 2 from 4 to 8192. (b) The precision as a function of  $\alpha$ .

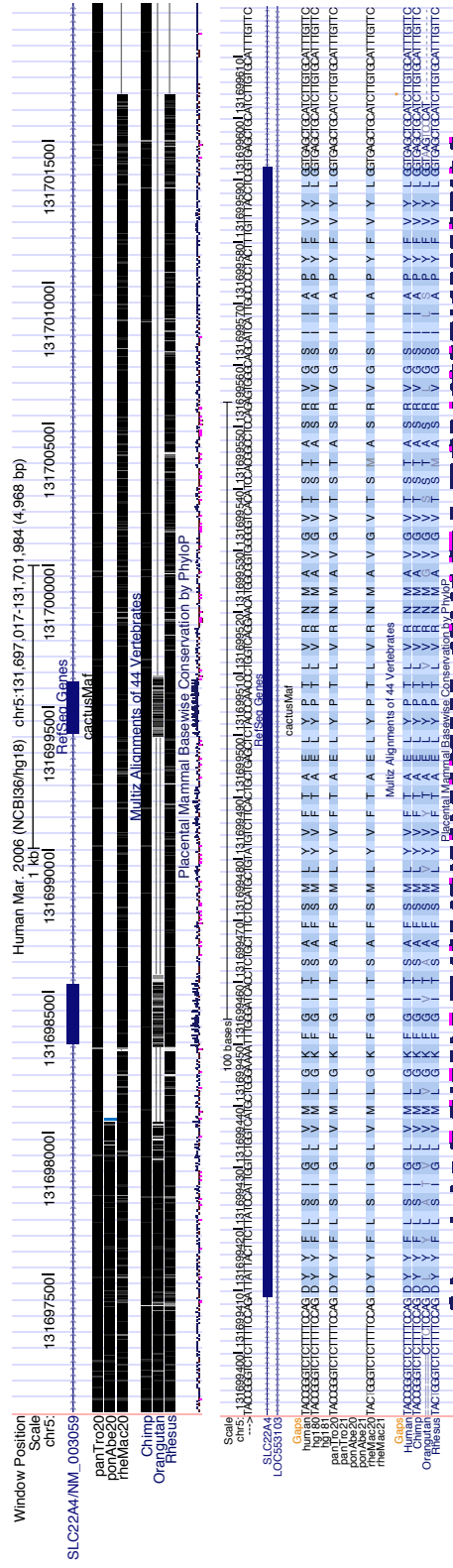


Figure 4: An example in which two exons in the SLC22A4 gene are conserved in the Multiz alignment, but non-conserved in the Cactus alignment due to a failure to align an orangutan sequence. Top, the exons in a 5 kilobase window. Bottom, a translated alignment view of the second missing exon. Multiz has aligned a sequence with 15 nonsynonymous mutations (in the first exon there are 17 such mutations). In contrast in this exon the rhesus sequence has just 1 nonsynonymous mutation, suggesting strongly that the aligned Orangutan sequence is not the true ortholog. Visualisation generated using the UCSC browser [Fujita et al., 2011].



8







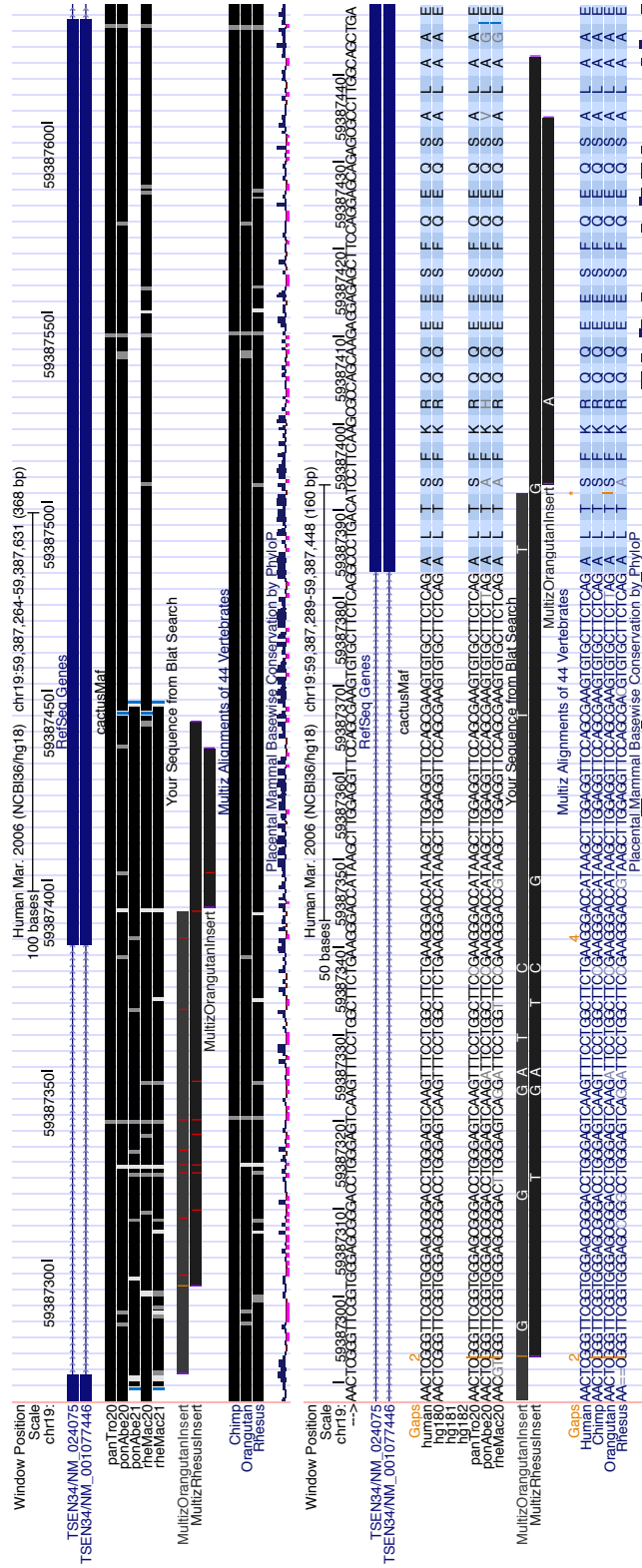


Figure 8: An example in which an exon is conserved in the Cactus alignment but non-conserved in the Multiz alignment, due to an apparent indel that actually arose by a tandem duplication. The first part of the exon is tandemly duplicated in both rhesus and orangutan. In the orangutan Multiz aligns the prefix of the first (5 prime) duplicated copy and the suffix of the second (3 prime) copy, creating an apparent 192 base pair insertion in the exon, while in Rhesus it correctly aligns only the second copy to the human exon. The cactus alignment contains both copies of the tandem duplication in both species. The Blat alignment shows the homology of the sequences unaligned by Multiz to the human sequence. Visualisation generated using the UCSC browser [Fujita et al., 2011].

Evolver Mammals Simulation

Program	Class	Human-Dog						Human-Mouse					
		Burn-In		No Burn-In		F1		Burn-In		No Burn-In		F1	
		R.	P.	R.	P.			R.	P.	R.	P.		
Cactus	All	<b>.585</b>	.946	<b>.831</b>	.946	<b>.885</b>	<b>.522</b>	<b>.911</b>	<b>.740</b>	<b>.911</b>	<b>.817</b>		
	No repeats	<b>.585</b>	.946	<b>.831</b>	.946	<b>.885</b>	<b>.522</b>	<b>.911</b>	<b>.740</b>	<b>.911</b>	<b>.817</b>		
	Repeats	<b>.669</b>	<b>.780</b>	<b>.726</b>	<b>.780</b>	<b>.752</b>	<b>.574</b>	.796	<b>.641</b>	.798	<b>.710</b>		
	Near	.646	.976	<b>.867</b>	.975	<b>.918</b>	<b>.593</b>	.959	<b>.794</b>	.958	<b>.869</b>		
Cactus(N.T.)	All	.520	<b>.947</b>	.739	<b>.947</b>	.830	.459	.910	.651	.910	.759		
	No repeats	.520	<b>.947</b>	.739	<b>.947</b>	.830	.459	.910	.651	.910	.759		
	Repeats	.598	.776	.650	.776	.707	.496	<b>.821</b>	.547	<b>.821</b>	.656		
	Near	.574	<b>.977</b>	.771	<b>.976</b>	.862	.522	<b>.960</b>	.699	<b>.959</b>	.809		
Pecan	All	.388	.920	.551	.919	.689	.357	.864	.505	.863	.638		
	No repeats	.388	.920	.551	.919	.689	.357	.864	.505	.863	.638		
	Repeats	.488	.721	.530	.721	.611	.419	.771	.462	.771	.578		
	Near	.429	.954	.576	.952	.719	.410	.922	.548	.919	.687		
TBA	All	.470	.836	.667	.835	.742	.459	.794	.650	.793	.715		
	No repeats	.470	.836	.667	.835	.742	.459	.794	.650	.793	.715		
	Repeats	.354	.517	.376	.506	.435	.341	.579	.376	.579	.456		
	Near	.535	.889	.717	.886	.793	.532	.857	.711	.854	.777		
TBA(ENCODE)	All	.483	.790	.684	.788	.733	.463	.794	.655	.792	.718		
	No repeats	.483	.790	.685	.788	.733	.463	.794	.655	.792	.718		
	Repeats	.362	.447	.393	.447	.418	.341	.571	.376	.571	.454		
	Near	.551	.841	.736	.837	.785	.536	.857	.716	.853	.780		
Multiz	All	.574	.851	.807	.843	.828	.470	.793	.664	.789	.723		
	No repeats	.574	.851	.807	.843	.828	.471	.793	.664	.789	.723		
	Repeats	.567	.632	.598	.614	.614	.349	.600	.385	.600	.469		
	Near	<b>.651</b>	<b>.903</b>	<b>.866</b>	.893	<b>.884</b>	<b>.545</b>	.857	<b>.727</b>	.851	<b>.786</b>		
pMauve	All	.387	.659	.549	.658	.599	.151	.674	.214	.673	.324		
	No repeats	.386	.659	.549	.658	.599	.151	.674	.214	.673	.324		
	Repeats	.433	.529	.470	.529	.498	.155	.556	.171	.556	.261		
	Near	.431	.689	.579	.688	.629	.173	.722	.232	.720	.352		

Table 1: Comparisons using the Evolver Mammals dataset. Format of the table is the same as in Table 1 of the main paper.

Evolver Mammals Simulation

Program	Class	Human-Cow						Mouse-Rat					
		Burn-In		No Burn-In		F1		Burn-In		No Burn-In		F1	
		R.	P.	R.	P.			R.	P.	R.	P.		
Cactus	All	<b>.576</b>	.939	<b>.818</b>	<b>.939</b>	<b>.874</b>	<b>.590</b>	.961	<b>.787</b>	.963	<b>.865</b>		
	No repeats	<b>.576</b>	.939	<b>.818</b>	<b>.939</b>	<b>.874</b>	<b>.590</b>	.961	<b>.787</b>	.963	<b>.865</b>		
	Repeats	<b>.596</b>	.730	<b>.675</b>	.730	<b>.701</b>	<b>.694</b>	.861	<b>.724</b>	.874	<b>.786</b>		
	Near	<b>.638</b>	<b>.972</b>	<b>.857</b>	<b>.971</b>	<b>.911</b>	<b>.643</b>	.972	<b>.817</b>	.973	<b>.887</b>		
Cactus(N.T.)	All	.502	<b>.940</b>	.713	<b>.939</b>	.811	.554	<b>.964</b>	.739	<b>.966</b>	.837		
	No repeats	.502	<b>.940</b>	.713	<b>.939</b>	.811	.554	<b>.964</b>	.739	<b>.966</b>	.837		
	Repeats	.529	<b>.774</b>	.600	<b>.774</b>	.676	.638	.856	.661	.870	.746		
	Near	.557	<b>.972</b>	.748	<b>.971</b>	.845	.603	<b>.973</b>	.766	<b>.975</b>	.857		
Pecan	All	.458	.911	.650	.910	.759	.486	.959	.648	.961	.774		
	No repeats	.458	.911	.650	.910	.759	.486	.959	.648	.961	.774		
	Repeats	.478	.691	.542	.691	.607	.587	<b>.878</b>	.609	<b>.893</b>	.719		
	Near	.510	.948	.685	.946	.795	.528	.968	.671	.970	.793		
TBA	All	.481	.840	.682	.838	.753	.573	.944	.765	.946	.845		
	No repeats	.481	.840	.682	.838	.753	.573	.944	.765	.946	.845		
	Repeats	.331	.529	.375	.529	.439	.663	.850	.687	.863	.760		
	Near	.550	.896	.737	.893	.809	.624	.955	.793	.957	.867		
TBA(ENCODE)	All	.454	.798	.643	.795	.712	.573	.944	.765	.946	.845		
	No repeats	.454	.799	.643	.795	.712	.573	.944	.765	.946	.845		
	Repeats	.346	.511	.392	.511	.443	.663	.850	.687	.863	.760		
	Near	.521	.853	.695	.848	.766	.624	.955	.793	.957	.867		
Multiz	All	.553	.842	.780	.837	.810	.435	.915	.580	.916	.710		
	No repeats	.553	.842	.780	.837	.810	.435	.915	.580	.916	.710		
	Repeats	.471	.587	.533	.587	.559	.179	.660	.193	.698	.298		
	Near	<b>.631</b>	.897	<b>.841</b>	.890	.868	.476	.927	.602	.927	.730		
pMauve	All	.344	.602	.489	.601	.539	.335	.571	.446	.572	.501		
	No repeats	.344	.602	.489	.601	.539	.335	.571	.446	.572	.501		
	Repeats	.331	.388	.375	.388	.381	.429	.519	.448	.531	.481		
	Near	.386	.632	.519	.631	.570	.366	.580	.465	.581	.516		

Table 2: Comparisons using the Evolver Mammals dataset. Format of the table is the same as in Table 1 of the main paper.

Evolver Mammals Simulation

Program	Class	Mouse-Dog						Mouse-Cow					
		Burn-In		No Burn-In		F1		Burn-In		No Burn-In		F1	
		R.	P.	R.	P.			R.	P.	R.	P.		
Cactus	All	.502	.898	.712	.897	.794	.494	.894	.894	.702	.892	.787	.787
	No repeats	.502	.898	.712	.897	.794	.494	.894	.894	.702	.892	.787	.787
	Repeats	.602	.725	.610	.725	.662	.469	.716	.525	.716	.716	.606	.606
	Near	.575	.953	.770	.950	.852	.568	.951	.762	.948	.846	.846	.846
Cactus(N.T.)	All	.425	.899	.603	.898	.722	.408	.896	.580	.895	.704	.704	.704
	No repeats	.425	.899	.603	.898	.722	.408	.896	.580	.895	.704	.704	.704
	Repeats	.506	.700	.512	.700	.592	.389	.746	.436	.733	.550	.550	.550
	Near	.487	.955	.652	.952	.775	.469	.953	.629	.951	.758	.758	.758
Pecan	All	.301	.841	.426	.839	.566	.349	.830	.495	.828	.620	.620	.620
	No repeats	.301	.841	.426	.839	.566	.349	.830	.495	.828	.620	.620	.620
	Repeats	.434	.632	.439	.632	.518	.310	.574	.347	.574	.432	.432	.432
	Near	.349	.905	.467	.902	.616	.406	.897	.544	.893	.677	.677	.677
TBA	All	.426	.753	.603	.751	.670	.434	.753	.615	.751	.677	.677	.677
	No repeats	.426	.753	.604	.751	.670	.434	.753	.615	.751	.677	.677	.677
	Repeats	.313	.456	.317	.456	.374	.327	.536	.366	.536	.435	.435	.435
	Near	.499	.819	.667	.815	.735	.511	.823	.683	.819	.746	.746	.746
TBA(ENCODE)	All	.450	.722	.638	.720	.677	.429	.733	.609	.731	.665	.665	.665
	No repeats	.450	.722	.638	.720	.677	.429	.733	.609	.731	.666	.666	.666
	Repeats	.398	.398	.402	.398	.400	.319	.500	.356	.500	.416	.416	.416
	Near	.528	.787	.706	.783	.744	.507	.803	.677	.798	.734	.734	.734
Multiz	All	.417	.770	.591	.768	.669	.409	.760	.579	.757	.657	.657	.657
	No repeats	.417	.770	.591	.769	.669	.409	.760	.579	.757	.658	.658	.658
	Repeats	.301	.463	.305	.463	.368	.248	.491	.277	.491	.354	.354	.354
	Near	.489	.839	.654	.835	.735	.482	.831	.643	.826	.725	.725	.725
pMauve	All	.160	.732	.226	.730	.346	.143	.669	.202	.667	.311	.311	.311
	No repeats	.160	.732	.226	.731	.346	.143	.669	.202	.667	.311	.311	.311
	Repeats	.120	.556	.122	.556	.200	.080	.346	.089	.346	.142	.142	.142
	Near	.185	.792	.249	.789	.378	.166	.725	.222	.722	.340	.340	.340

Table 3: Comparisons using the Evolver Mammals dataset. Format of the table is the same as in Table 1 of the main paper.

Evolver Mammals Simulation

Program	Class	Rat-Human				Rat-Dog					
		Burn-In		No Burn-In		Burn-In		No Burn-In			
		R.	P.	R.	P.	R.	P.	R.	P.		
Cactus	All	.522	.911	.738	.911	.815	.502	.898	.710	.897	.793
	No repeats	.522	.911	.738	.911	.815	.502	.898	.710	.897	.793
	Repeats	.617	.791	.677	.800	.729	.590	.737	.615	.737	.670
	Near	.592	.960	.791	.959	.867	.575	.953	.768	.951	.850
Cactus(N.T.)	All	.458	.910	.646	.910	.756	.423	.900	.599	.899	.719
	No repeats	.458	.910	.646	.910	.756	.423	.900	.599	.899	.719
	Repeats	.532	.815	.585	.826	.681	.490	.700	.510	.700	.590
	Near	.520	.960	.694	.959	.805	.485	.955	.648	.954	.772
Pecan	All	.304	.866	.429	.864	.574	.252	.843	.357	.841	.501
	No repeats	.304	.866	.429	.864	.574	.252	.843	.357	.841	.501
	Repeats	.355	.735	.385	.735	.505	.390	.684	.406	.684	.510
	Near	.349	.921	.464	.919	.618	.292	.906	.390	.903	.545
TBA	All	.453	.791	.639	.790	.707	.422	.750	.596	.747	.664
	No repeats	.453	.791	.639	.790	.707	.422	.750	.596	.748	.664
	Repeats	.390	.561	.431	.571	.487	.290	.460	.302	.460	.365
	Near	.524	.854	.698	.851	.768	.495	.816	.659	.812	.729
TBA(ENCODE)	All	.459	.790	.646	.788	.711	.450	.718	.635	.716	.674
	No repeats	.459	.790	.646	.788	.711	.450	.718	.635	.716	.674
	Repeats	.390	.556	.431	.566	.485	.390	.402	.406	.402	.404
	Near	.532	.853	.707	.849	.773	.528	.784	.703	.780	.741
Multiz	All	.467	.780	.656	.777	.713	.415	.760	.586	.758	.662
	No repeats	.467	.780	.656	.777	.713	.415	.760	.586	.758	.662
	Repeats	.411	.547	.454	.557	.496	.280	.431	.292	.431	.348
	Near	.541	.844	.719	.839	.776	.488	.829	.649	.825	.728
pMauve	All	.107	.469	.151	.468	.229	.118	.536	.167	.535	.255
	No repeats	.107	.469	.151	.468	.229	.118	.536	.167	.535	.255
	Repeats	.092	.283	.100	.283	.148	.070	.280	.073	.280	.116
	Near	.123	.501	.164	.500	.248	.137	.579	.183	.578	.278

Table 4: Comparisons using the Evolver Mammals dataset. Format of the table is the same as in Table 1 of the main paper.

Evolver Mammals Simulation						
Program	Class	Rat-Cow				
		Burn-In		No Burn-In		F1
		R.	P.	R.	P.	
Cactus	All	<b>.494</b>	.893	<b>.699</b>	.892	<b>.784</b>
	No repeats	<b>.494</b>	.893	<b>.699</b>	.892	<b>.784</b>
	Repeats	<b>.496</b>	.722	<b>.564</b>	.722	<b>.633</b>
	Near	<b>.568</b>	.950	<b>.759</b>	.948	<b>.844</b>
Cactus(N.T.)	All	.408	<b>.897</b>	.577	<b>.896</b>	.702
	No repeats	.408	<b>.897</b>	.577	<b>.896</b>	.702
	Repeats	.400	<b>.754</b>	.455	<b>.742</b>	.568
	Near	.469	<b>.954</b>	.626	<b>.952</b>	.756
Pecan	All	.295	.826	.417	.825	.554
	No repeats	.295	.826	.417	.825	.554
	Repeats	.287	.579	.317	.561	.410
	Near	.344	.892	.458	.889	.605
TBA	All	.430	.750	.607	.748	.671
	No repeats	.430	.750	.607	.748	.671
	Repeats	.296	.507	.337	.507	.405
	Near	.507	.820	.674	.815	.740
TBA(ENCODE)	All	.430	.730	.607	.727	.663
	No repeats	.430	.730	.607	.727	.663
	Repeats	.304	.461	.347	.461	.395
	Near	.508	.799	.675	.794	.732
Multiz	All	.407	.747	.574	.745	.649
	No repeats	.407	.748	.574	.745	.649
	Repeats	.252	.453	.287	.453	.352
	Near	.480	.819	.638	.814	.717
pMauve	All	.100	.469	.142	.468	.218
	No repeats	.100	.469	.142	.468	.218
	Repeats	.061	.189	.069	.189	.101
	Near	.117	.508	.156	.506	.239

Table 5: Comparisons using the Evolver Mammals dataset. Format of the table is the same as in Table 1 of the main paper.



Evolver Primates Simulation

Program	Class	Human-Chimp						Human-Gorilla					
		Burn-In		No Burn-In		F1		Burn-In		No Burn-In		F1	
		R.	P.	R.	P.	R.	P.	R.	P.	R.	P.	R.	P.
Cactus	All	.709	.993	.990	.981	.992	.710	.993	.991	.981	.992	.992	.992
	No repeats	.709	.993	.990	.981	.992	.710	.993	.991	.981	.992	.992	.992
	Repeats	.853	.938	.961	.938	.949	.821	.934	.948	.934	.941	.941	.941
	Near	.748	.993	.992	.981	.993	.749	.993	.993	.981	.993	.993	.993
Cactus(N.T.)	All	.698	<b>.999</b>	.987	.998	<b>.993</b>	.699	.998	.987	.998	<b>.993</b>	<b>.993</b>	<b>.993</b>
	No repeats	.698	<b>.999</b>	.987	.998	<b>.993</b>	.699	.998	.987	.998	<b>.993</b>	<b>.993</b>	<b>.993</b>
	Repeats	.853	<b>.999</b>	.961	<b>.999</b>	<b>.980</b>	.821	.977	.948	.977	<b>.962</b>	<b>.962</b>	<b>.962</b>
	Near	.736	<b>.999</b>	.989	.998	<b>.994</b>	.737	<b>.999</b>	.988	.998	<b>.994</b>	<b>.994</b>	<b>.994</b>
Pecan	All	.696	<b>.999</b>	.985	<b>.999</b>	.992	.698	<b>.999</b>	.987	<b>.999</b>	<b>.993</b>	<b>.993</b>	<b>.993</b>
	No repeats	.696	<b>.999</b>	.985	<b>.999</b>	.992	.698	<b>.999</b>	.987	<b>.999</b>	<b>.993</b>	<b>.993</b>	<b>.993</b>
	Repeats	.846	.992	.953	.992	.972	.814	<b>.984</b>	.941	<b>.984</b>	<b>.962</b>	<b>.962</b>	<b>.962</b>
	Near	.734	<b>.999</b>	.987	<b>.999</b>	.993	.736	<b>.999</b>	.988	<b>.999</b>	<b>.994</b>	<b>.994</b>	<b>.994</b>
TBA	All	.698	.997	.987	.997	.992	.696	.998	.983	.998	.991	.991	.991
	No repeats	.698	.997	.987	.997	.992	.696	.998	.983	.998	.991	.991	.991
	Repeats	.853	.992	.961	.992	.976	.814	<b>.984</b>	.941	<b>.984</b>	<b>.962</b>	<b>.962</b>	<b>.962</b>
	Near	.736	.997	.989	.997	.993	.734	<b>.999</b>	.985	<b>.999</b>	.992	.992	.992
TBA(ENCODE)	All	<b>.722</b>	.950	<b>.993</b>	.923	.971	.716	.979	.987	.954	.983	.983	.983
	No repeats	<b>.722</b>	.950	<b>.993</b>	.923	.971	.716	.979	.987	.954	.983	.983	.983
	Repeats	<b>.867</b>	.879	<b>.976</b>	.879	.925	.821	.914	.948	.914	.931	.931	.931
	Near	<b>.762</b>	.950	<b>.995</b>	.923	.972	.756	.979	.989	.954	.984	.984	.984
Multiz	All	<b>.722</b>	.950	<b>.993</b>	.923	.971	<b>.723</b>	.955	<b>.993</b>	.928	.974	.974	.974
	No repeats	<b>.722</b>	.950	<b>.993</b>	.923	.971	<b>.723</b>	.955	<b>.993</b>	.928	.974	.974	.974
	Repeats	<b>.867</b>	.879	<b>.976</b>	.879	.925	<b>.827</b>	.908	<b>.956</b>	.908	.931	.931	.931
	Near	<b>.762</b>	.950	<b>.995</b>	.923	.972	<b>.763</b>	.956	<b>.995</b>	.929	.975	.975	.975
pMauve	All	.693	<b>.999</b>	.981	<b>.999</b>	.990	.690	.998	.975	.998	.987	.987	.987
	No repeats	.693	<b>.999</b>	.981	<b>.999</b>	.990	.690	.998	.975	.998	.987	.987	.987
	Repeats	.846	<b>.999</b>	.953	<b>.999</b>	.976	.814	<b>.984</b>	.941	<b>.984</b>	<b>.962</b>	<b>.962</b>	<b>.962</b>
	Near	.731	<b>.999</b>	.983	<b>.999</b>	.991	.728	<b>.999</b>	.977	<b>.999</b>	.988	.988	.988

Table 6: Comparisons using the Evolver Primates dataset. Format of the table is the same as in Table 1 of the main paper.

Evolver Primates Simulation

Program	Class	Human-Orang						Chimp-Gorilla					
		Burn-In		No Burn-In		F1		Burn-In		No Burn-In		F1	
		R.	P.	R.	P.			R.	P.	R.	P.		
Cactus	All	.707	.992	.991	.980	<b>.992</b>		<b>.708</b>	.993	<b>.989</b>	.980	<b>.991</b>	
	No repeats	.707	.992	.991	.980	.992		<b>.708</b>	.993	<b>.989</b>	.980	<b>.991</b>	
	Repeats	.865	.917	.946	.917	.931		<b>.808</b>	.920	<b>.940</b>	.920	.930	
	Near	.746	.993	<b>.994</b>	.981	.993		<b>.747</b>	.993	<b>.991</b>	.981	<b>.992</b>	
Cactus(N.T.)	All	.695	<b>.998</b>	.987	<b>.997</b>	<b>.992</b>		.697	.998	.985	.998	<b>.991</b>	
	No repeats	.695	<b>.998</b>	.987	<b>.997</b>	<b>.993</b>		.697	.998	.985	.998	<b>.991</b>	
	Repeats	.879	.969	.961	.969	<b>.965</b>		<b>.808</b>	.977	<b>.940</b>	.977	.958	
	Near	.734	<b>.999</b>	.989	<b>.998</b>	<b>.994</b>		.735	<b>.999</b>	.986	.998	<b>.992</b>	
Pecan	All	.687	.997	.976	<b>.997</b>	.987		.695	<b>.999</b>	.983	<b>.999</b>	<b>.991</b>	
	No repeats	.687	.997	.976	<b>.997</b>	.987		.695	<b>.999</b>	.983	<b>.999</b>	<b>.991</b>	
	Repeats	.837	.967	.915	.967	.940		<b>.808</b>	<b>.992</b>	<b>.940</b>	<b>.992</b>	<b>.966</b>	
	Near	.725	.998	.978	<b>.998</b>	.988		.733	<b>.999</b>	.984	<b>.999</b>	<b>.992</b>	
TBA	All	.694	.997	.986	<b>.997</b>	.991		.695	.996	.983	.996	.990	
	No repeats	.694	.997	.986	<b>.997</b>	.991		.695	.996	.983	.996	.990	
	Repeats	.872	<b>.976</b>	.953	<b>.976</b>	<b>.965</b>		<b>.808</b>	.977	<b>.940</b>	.977	.958	
	Near	.733	.998	.988	<b>.998</b>	.993		.733	.996	.985	.996	.991	
TBA(ENCODE)	All	.713	.979	.988	.954	.983		.695	.995	.984	.995	.989	
	No repeats	.713	.979	.988	.954	.983		.695	.995	.984	.995	.989	
	Repeats	.872	.918	.953	.918	.935		<b>.808</b>	.977	<b>.940</b>	.977	.958	
	Near	.754	.980	.990	.955	.985		.733	.996	.985	.996	.991	
Multiz	All	<b>.719</b>	.953	<b>.992</b>	.926	.972		.695	.995	.983	.995	.989	
	No repeats	<b>.719</b>	.953	<b>.992</b>	.926	.972		.695	.995	.983	.995	.989	
	Repeats	<b>.887</b>	.912	<b>.969</b>	.912	.940		<b>.808</b>	.977	<b>.940</b>	.977	.958	
	Near	<b>.760</b>	.955	<b>.994</b>	.927	.974		.733	.996	.985	.996	.991	
pMauve	All	.678	.989	.964	.989	.977		.689	.998	.974	.998	.986	
	No repeats	.678	.989	.964	.989	.977		.689	.998	.974	.998	.986	
	Repeats	.830	.951	.907	.951	.929		<b>.808</b>	.984	<b>.940</b>	.984	.962	
	Near	.716	.991	.966	.991	.978		.726	<b>.999</b>	.976	<b>.999</b>	.987	

Table 7: Comparisons using the Evolver Primates dataset. Format of the table is the same as in Table 1 of the main paper.

Evolver Primates Simulation

Program	Class	Chimp-Orang						Gorilla-Orang					
		Burn-In		No Burn-In		F1		Burn-In		No Burn-In		F1	
		R.	P.	R.	P.			R.	P.	R.	P.		
Cactus	All	<b>.705</b>	.992	<b>.990</b>	.980	<b>.991</b>		<b>.707</b>	.992	<b>.991</b>	.980	<b>.992</b>	
	No repeats	<b>.705</b>	.992	<b>.990</b>	.980	<b>.991</b>		<b>.707</b>	.992	<b>.991</b>	.980	<b>.992</b>	
	Repeats	.846	.910	.931	.910	.920		.840	.926	.955	.926	.940	
	Near	<b>.745</b>	.993	<b>.992</b>	.981	<b>.993</b>		<b>.746</b>	.993	<b>.993</b>	.981	<b>.993</b>	
Cactus(N.T.)	All	.694	<b>.998</b>	.985	<b>.997</b>	<b>.991</b>		.694	<b>.998</b>	.985	<b>.997</b>	.991	
	No repeats	.694	<b>.998</b>	.985	<b>.997</b>	<b>.991</b>		.694	<b>.998</b>	.985	<b>.997</b>	.991	
	Repeats	<b>.860</b>	.969	<b>.946</b>	.969	<b>.957</b>		<b>.853</b>	.977	<b>.970</b>	.977	<b>.973</b>	
	Near	.733	<b>.998</b>	.988	<b>.998</b>	<b>.993</b>		.733	<b>.999</b>	.987	<b>.998</b>	<b>.993</b>	
Pecan	All	.684	.997	.972	<b>.997</b>	.985		.686	.997	.974	<b>.997</b>	.985	
	No repeats	.684	.997	.972	<b>.997</b>	.985		.686	.997	.974	<b>.997</b>	.985	
	Repeats	.818	<b>.975</b>	.900	<b>.975</b>	.936		.820	.984	.932	.984	.957	
	Near	.723	<b>.998</b>	.975	<b>.998</b>	.986		.724	.998	.976	<b>.998</b>	.987	
TBA	All	.692	.995	.984	.995	.989		.692	.996	.983	.996	.989	
	No repeats	.692	.995	.984	.995	.989		.692	.996	.983	.996	.989	
	Repeats	.853	.968	.938	.968	.953		.833	<b>.992</b>	.947	<b>.992</b>	.969	
	Near	.731	.996	.986	.996	.991		.731	.997	.986	.997	.991	
TBA(ENCODE)	All	.693	.994	.984	.994	.989		.693	.996	.984	.996	.990	
	No repeats	.693	.994	.984	.994	.989		.693	.996	.984	.996	.990	
	Repeats	<b>.860</b>	.961	<b>.946</b>	.961	.953		.840	.984	.955	.984	.969	
	Near	.732	.995	.987	.995	.991		.732	.997	.986	.997	.991	
Multiz	All	.693	.994	.984	.994	.989		.693	.996	.984	.996	.990	
	No repeats	.693	.994	.984	.994	.989		.693	.996	.984	.996	.990	
	Repeats	.853	.961	.938	.961	.949		.847	.984	.962	.984	<b>.973</b>	
	Near	.731	.996	.987	.996	.991		.732	.997	.987	.997	.992	
pMauve	All	.676	.989	.961	.989	.975		.677	.989	.961	.989	.975	
	No repeats	.676	.989	.961	.989	.975		.677	.989	.961	.989	.975	
	Repeats	.818	.944	.900	.944	.921		.820	.969	.932	.969	.950	
	Near	.714	.990	.964	.990	.977		.715	.990	.964	.990	.977	

Table 8: Comparisons using the Evolver Primates dataset. Format of the table is the same as in Table 1 of the main paper.

Primates					
Exon Id	Gene	Accession	Cactus	Multiz	Comment
1	TSEN34	NM_001077446	C	N	IE
2	KIR2DL3	NM_015868	C	N	IE
7	KIR2DL3	NM_015868	C	N	IE
7	KIR2DL1	NM_014218	C	N	IE
8	GART	NM_000819	C	N	IE
0	GART	NM_000819	C	N	IE
12	LILRB3	NM_001081450	C	N	NA
11	LILRB3	NM_001081450	C	N	NA
10	LILRB3	NM_001081450	C	N	NA
9	LILRB3	NM_001081450	C	N	NA
8	LILRB3	NM_001081450	C	N	NA
7	LILRB3	NM_001081450	C	N	NA
6	LILRB3	NM_001081450	C	N	NA
4	LILRB3	NM_001081450	C	N	NA
3	LILRB3	NM_001081450	C	N	NA
2	LILRB3	NM_001081450	C	N	NA
1	LILRB3	NM_001081450	C	N	NA
0	LILRB3	NM_001081450	C	N	NA
7	LILRA6	NM_024318	C	N	NA
12	LILRB5	NM_001081442	C	N	NA
11	LILRB5	NM_001081442	C	N	NA
10	LILRB5	NM_001081442	C	N	NA
8	LILRB5	NM_001081442	C	N	NA
7	LILRB5	NM_001081442	C	N	NA
6	LILRB5	NM_001081442	C	N	NA
4	LILRB5	NM_001081442	C	N	NA
3	LILRB5	NM_001081442	C	N	NA
2	LILRB5	NM_001081442	C	N	NA
1	LILRB5	NM_001081442	C	N	NA
0	LILRB5	NM_001081442	C	N	NA
12	LILRB2	NM_005874	C	N	NA
11	LILRB2	NM_005874	C	N	NA
10	LILRB2	NM_005874	C	N	NA
9	LILRB2	NM_005874	C	N	NA
8	LILRB2	NM_005874	C	N	NA
2	LILRB2	NM_005874	C	N	NA
1	LILRB2	NM_005874	C	N	NA
0	LILRB2	NM_005874	C	N	NA
7	C22orf42	NM_001010859	C	N	NA
5	C22orf42	NM_001010859	C	N	NA
4	C22orf42	NM_001010859	C	N	NA
3	C22orf42	NM_001010859	C	N	NA
2	C22orf42	NM_001010859	C	N	NA
1	C22orf42	NM_001010859	C	N	NA
0	C22orf42	NM_001010859	C	N	NA
0	TEX28	NM_001586	C	N	NA
0	OR52R1	NM_001005177	C	N	WH
0	OR51T1	NM_001004759	C	N	WH
0	OR51L1	NM_001004755	C	N	WH
0	OR52J3	NM_001001916	C	N	WH
0	OR52E2	NM_001005164	C	N	WH
1	HBD	NM_000519	C	N	WH
1	HBG1	NM_000559	C	N	WH
1	HBG2	NM_000184	C	N	WH
0	OR51B6	NM_001004750	C	N	WH
1	AXIN1	NM_003502	C	N	WH
4	LILRA6	NM_024318	C	N	WH
7	LILRB2	NM_005874	C	N	WH
6	LILRB2	NM_005874	C	N	WH
5	LILRB2	NM_005874	C	N	WH
4	LILRB2	NM_005874	C	N	WH
5	LILRA3	NM_006865	C	N	WH
2	LILRA3	NM_006865	C	N	WH
8	LAIR1	NM_002287	C	N	WH
0	LAIR1	NM_002287	C	N	WH
4	LILRA1	NM_006863	C	N	WH
4	LILRB4	NM_006847	C	N	WH
6	KIR2DL3	NM_015868	C	N	WH
6	KIR2DL4	NM_002255	C	N	WH
14	ITSN1	NM_003024	C	N	WH
3	RFPL2	NM_001098527	C	N	WH
1	RFPL2	NM_001098527	C	N	WH
1	CTAG1A	NM_139250	C	N	WH
1	CTAG1B	NM_001327	C	N	WH
3	FUNDC2	NM_023934	C	N	WH

Table 9: Exons that are conserved in Cactus, but non-conserved in Multiz. C: Conserved, N: Non-conserved, WH: Wrong Homolog, in which a non-conserved sequence is aligned in preference to a more closely homologous sequence in one or more species, IE: Indel Error, in which the alignment contains a pattern of indels in all the sequences from one or more species that prevent the sequences from being considered conserved, NA: Not Aligned, where no sequence is aligned in one or more species.

Primates					
Exon Id	Gene	Accession	Cactus	Multiz	Comment
0	OR51B2	NM_033180	N	C	NA
5	LILRA4	NM_012276	N	C	NA
4	LILRA4	NM_012276	N	C	NA
3	LILRA4	NM_012276	N	C	NA
2	LILRA4	NM_012276	N	C	NA
1	LILRA4	NM_012276	N	C	NA
0	LILRA4	NM_012276	N	C	NA
5	KIR2DL1	NM_014218	N	C	NA
6	SLC22A4	NM_003059	N	C	NA
7	SLC22A4	NM_003059	N	C	NA
5	LILRA6	NM_024318	N	C	WH

Table 10: Exons that are conserved in Multiz, but non-conserved in Cactus. Same format as Table 9.

Primates loci reconstruction dataset				
Region	Species	Chromosome	Start	End
ENm001	hg18	chr7	115597756	117475182
	panTro2	chr7	116358580	118271369
	ponAbe2	chr7_random	15766758	15905393
		chr7	112761783	114610423
ENm002	rheMac2	chr3	153523204	155364336
	hg18	chr5	131284313	132284313
	panTro2	chr5	133541207	134551980
	ponAbe2	chr5	133353386	134373402
ENm003	rheMac2	chr6	128348108	129327973
	hg18	chr11	115962315	116462315
	panTro2	chr11	115397026	115908101
	ponAbe2	chr11	113509259	114016230
ENm004	rheMac2	chr14	114993392	115485494
	hg18	chr22	30133953	31833953
	panTro2	chr22	30294491	32011499
	ponAbe2	chr22	26374071	28234221
ENm005	rheMac2	chr10	75300555	77028038
		chr10	88219525	88274424
	hg18	chr21	32668236	34364221
	panTro2	chr21	32122073	33817412
ENm006	ponAbe2	chr21	33317601	35079767
	rheMac2	chr19	8303957	8346701
		chr3	12498479	14135229
	hg18	chrX	152767491	154063081
ENm007	panTro2	chrX	153291711	154485343
	ponAbe2	chrX	154044151	155368060
	rheMac2	chrX	151961488	153195575
		chr19	59023584	60024460
ENm008	hg18	chr19	59513067	60564353
	panTro2	chr19	55606585	56682663
	rheMac2	chr19	59953290	60901251
		chr16	45	500000
ENm009	hg18	chr16	20539	520583
	panTro2	chr16	4664777	4675171
	ponAbe2	chrX	156177209	156188089
		chr16	10000	522714
ENm010	rheMac2	chr2b	21132250	21142279
		chr20	458	494348
	rheMac2	chr13	113700995	113708122
		chr15	61543	73215
ENm011	hg18	chr11	4730995	5732587
	panTro2	chr11	4579020	5592480
	ponAbe2	chr11	64509194	65984125
	rheMac2	chr14	67368393	68925291
ENm012	hg18	chr7	26924045	27424045
	panTro2	chr7	27260287	27768256
	ponAbe2	chr7	57012283	57540815
	rheMac2	chr22	98807544	99304888

Table 11: Input sequences for loci alignment. ‘Region’: name of the encode pilot region. ‘Species’: UCSC species assembly name. ‘Start’ and ‘End’: the locations of input sequences on the positive