Suppl. Table S1 Polymorphic di- and trinucleotide microsatellite repeats identified in little bush moa

Locus	Scaffold	Start (Repeat)	End (Repeat)	Start (Repeat + flank)	End (Repeat + flank)	REF Allele	ALT Allele
Dinucleotide repeats							
anoDid_di1	scaffold_423	460986	461001	460740	461181	$(AG)_8$	$(AG)_9$
anoDid_di2	scaffold_48	3097063	3097076	3096814	3097327	$(GT)_7$	$(GT)_6$
anoDid_di3	scaffold_50	996878	996899	996641	997129	$(AG)_{11}$	$(AG)_9$
anoDid_di4	scaffold_6	2351772	2351783	2351542	2352009	$(CT)_6$	$(CT)_5$
anoDid_di5	scaffold_60	2709321	2709346	2709103	2709574	$(GT)_{13}$	$(GT)_{12}$
anoDid_di6	scaffold_79	2063606	2063617	2063357	2063853	$(CT)_6$	$(CT)_7$
anoDid_di7	scaffold_81	3721848	3721869	3721619	3722094	$(GT)_{11}$	$(GT)_{13}$
anoDid_di8	scaffold_126	739220	739239	739065	739478	$(GT)_{10}$	$(GT)_{11}$
anoDid_di9	scaffold_135	108053	108068	107877	108297	$(CT)_8$	$(CT)_9$
anoDid_di10	scaffold_136	1740652	1740667	1740413	1740892	$(AC)_8$	$(AC)_7$
anoDid_di11	scaffold_158	2414894	2414905	2414744	2415137	$(AC)_6$	$(AC)_7$
anoDid_di12	scaffold_17	5603523	5603538	5603294	5603786	$(AC)_8$	$(AC)_9$
anoDid_di13	scaffold_18	221484	221499	221325	221741	$(AG)_8$	$(AG)_7$
anoDid_di14	scaffold_202	427665	427676	427435	427912	$(AC)_6$	$(AC)_5$
anoDid_di15	scaffold_21	4341625	4341636	4341376	4341869	$(CT)_6$	$(CT)_7$
anoDid_di16	scaffold_225	1492143	1492166	1491914	1492411	$(AC)_{12}$	$(AC)_{11}$
anoDid_di17	scaffold_26	3615418	3615431	3615208	3615669	$(AG)_7$	$(AG)_6$
anoDid_di18	scaffold_26	5268731	5268744	5268482	5268976	$(AC)_7$	$(AC)_9$
anoDid_di19	scaffold_105	1805021	1805038	1804919	1805278	$(AG)_9$	$(AG)_8$
anoDid_di20	scaffold_28	3056646	3056657	3056413	3056870	$(AG)_6$	$(AG)_7$
anoDid_di21	scaffold_30	1241411	1241430	1241184	1241551	$(GT)_{10}$	$(GT)_{11}$
anoDid_di22	scaffold_30	3716946	3716959	3716703	3717203	$(AT)_7$	$(AT)_6$
anoDid_di23	scaffold_107	1738948	1738959	1738700	1739207	$(CT)_6$	$(CT)_7$
anoDid_di24	scaffold_317	517104	517119	516868	517366	$(AC)_8$	$(AC)_6$
anoDid_di25	scaffold_368	665419	665444	665225	665692	$(GT)_{13}$	$(GT)_{14}$
anoDid_di26	scaffold_37	4246735	4246752	4246486	4246868	$(AG)_9$	$(AG)_8$
anoDid_di27	scaffold_40	4006807	4006820	4006562	4007061	$(AT)_7$	$(AT)_6$
anoDid_di28	scaffold_41	1770140	1770153	1769891	1770389	$(GT)_7$	$(GT)_6$
Trinucleotide repeats							
anoDid_tri1	scaffold_494	312050	312076	311833	312325	$(AAC)_9$	$(AAC)_8$
anoDid_tri2	scaffold_67	2212361	2212372	2212117	2212574	$(GCT)_4$	$(GCT)_5$
anoDid_tri3	scaffold_84	2470009	2470023	2469776	2470238	$(AGC)_5$	$(AGC)_6$
anoDid_tri4	scaffold_91	3565046	3565057	3564839	3565293	$(GTT)_4$	$(GTT)_5$
anoDid_tri5	scaffold_124	1297692	1297703	1297623	1297941	$(AGG)_4$	$(AGG)_3$
anoDid_tri6	scaffold_151	866499	866519	866256	866633	$(GTT)_7$	$(GTT)_9$
anoDid_tri7	scaffold_219	836816	836839	836581	836974	$(CCT)_8$	$(CCT)_9$
anoDid_tri8	scaffold_239	709585	709614	709338	709856	$(CGG)_{11}$ †	$(CGG)_8$
anoDid_tri9	scaffold_309	618397	618408	618175	618605	$(AAG)_4$	$(AAG)_5$
anoDid_tri10	scaffold_320	505190	505201	504969	505424	$(GCT)_4$	$(GCT)_5$
anoDid_tri11	scaffold_351	134675	134689	134456	134895	$(GCT)_5$	$(GCT)_4$
anoDid_tri12	scaffold_36	1888435	1888449	1888200	1888638	$(GCT)_6$ †	$(GCT)_7$

† differs from genome assembly by +1 repeat unit following indel realignment

Suppl. Table S2 Moa variants with PROVEAN score < -5 compared to the emu reference

Gene	Variant[1]	Alignment Position (AA)	PROVEAN (Emu-Moa)	Shared[2]	Alternative[3] (PROVEAN)	DoC[4]	Moa alleles
DCHS1	P2594L	2654	-7.364	No	S (-5.572)	4X (3X)	
	Q3065del	3126	-5.600	Yes	P (-1.429)	8X (6X)	
	H3199L	3267	-6.496	Yes	n/a	2X (1X)	
DVL1	N206Y	263	-5.080	No	n/a	3X (2X)	Tyr Y (TAC, 1X DoC) Asn N (AAC, 1X DoC)
DYNC2H1	P789A	793	-6.517	No	n/a	2X	
	M951T	955	-5.099	No	n/a	2X	
	C2494Y	2498	-6.584	No	R (-5.204)	6X	
EVC	H544R	603	-5.103	Yes	Q (-4.455)	4X	
	D703G	762	-5.620	Yes	E (-3.157) N (-4.149)	7X (6X)	
FAT1	E2568V	2570	-5.223	No	G (-4.742) K (-2.900) Q (-2.032)	1X	
	K2919I	2921	-6.160	No	R (-2.203)	4X	
	P3411R	3413	-6.287	Yes	A (-4.538) S (-5.030)	6X (5X)	
	H4379P	4385	-5.295	No	N (-3.224) Y (-2.563)	3X (2X)	
GLI2	A951_L952insL	982	-7.986	No	n/a	6X	
HOXD8	E124G	258	-6.488	No	n/a	5X (4X)	
KIF7	L459Q	641	-5.210	Yes	M (-1.685)	8X (6X)	
OFD1	E190A	190	-5.926	No	K (-3.947) Q (-2.958)	7X (6X)	
	R896del	917	-6.903	Yes	K (-1.404) M (-3.376) S (-3.318)	6X (5X)	
TALPID3	H766Q	856	-7.407	Yes	n/a	3X	Gln Q (CAG, 2X DoC) His H (CAC, 1X DoC)
	P817A	907	-7.738	Yes	T (-7.691)	5X	
	P1163A	1281	-7.066	Yes	n/a	19X (7X)	
	P1218A	1336	-6.453	Yes	H (-6.836) L (-7.451) S (-6.270) T (-6.366)	7X (6X)	
	P1229R	1347	-6.538	No	A (-6.168)	7X (5X)	
	P1379L	1501	-6.853	No	A (-5.107) S (-5.243)	8X (7X)	

[1]Variants are listed using HGVS (Human Genome Variation Society) notation. For example, P2594L indicates P at position 2594 in the emu reference amino acid sequence is replaced by L in moa.

[2]Indicates whether moa amino acid replacement is shared by other birds in alignment.

[3]Indicates alternative amino acid replacement present in other birds in alignment, with PROVEAN score relative to the emu reference in brackets.

[4]Where different, moa raw depth of coverage (DoC) is followed by DoC counting overlapping members of a single read pair as 1X (in brackets).

Suppl. Table S3 Moa variants with PROVEAN score < -5 compared to a moa-tinamou ancestral reference sequence. Variants differing from those using an emu reference are shown in bold.

Gene	Variant[1]	Alignment Position (AA)	PROVEAN (Anc-Moa)	Emu reference equivalent	Shared[2]	Alternative[3] (PROVEAN)	DoC[4]	Moa alleles
DCHS1	**P1989L**	**1989**	**-5.116**	**P1945L (-4.799)**	**No**	**S (-2.920)**	**6X (5X)**	
	P2654L	2654	-7.344	P2594L (-7.364)	No	S (-5.572)	4X (3X)	
	Q3126del	3126	-5.064	Q3065del (-5.600)	Yes	P (-1.429)	8X (6X)	
DVL1	N263Y	263	-5.069	N206Y (-5.080)	No	n/a	3X (2X)	Tyr Y (TAC, 1X DoC) Asn N (AAC, 1X DoC)
DYNC2H1	P793A	793	-6.518	P789A (-6.514)	No	n/a	2X	
	M955T	955	-5.233	M951T (-5.099)	No	n/a	2X	
	C2498Y	2498	-6.980	C2494Y (-6.584)	No	R (-5.204)	6X	
EVC	H603R	603	-5.087	H544R (-5.103)	Yes	Q (-4.455)	4X	
	D762G	762	-5.516	D703G (-5.620)	Yes	E (-3.157) N (-4.149)	7X (6X)	
FAT1	E2570V	2570	-5.409	E2568V (-5.223)	No	G (-4.742) K (-2.900) Q (-2.032)	1X	
	K2921I	2921	-6.226	K2919I (-6.160)	No	R (-2.203)	4X	
	H4385P	4385	-5.216	H4379P (-5.295)	No	N (-3.224) Y (-2.563)	3X (2X)	
HOXD8	E258G	258	-6.549	E124G (-6.488)	No	n/a	5X (4X)	
OFD1	E190A	190	-5.961	E190A (-5.926)	No	K (-3.947) Q (-2.958)	7X (6X)	
TALPID3	H856Q	856	-7.335	H766Q (-7.407)	Yes	n/a	3X	Gln Q (CAG, 2X DoC) His H (CAC, 1X DoC)
	M1127T	**1127**	**-5.056**	**M1028T (-4.731)**	**No**	**I (-3.099) L (-2.436) V (-3.059)**	**8X**	**Thr T (ACG, 7X DoC) Thr T (ACA, 1X DoC)**
	P1281A	1281	-6.827	P1163A (-7.066)	Yes	n/a	19X (7X)	
	P1336A	1336	-6.456	P1218A (-6.453)	Yes	H (-6.836)	7X (6X)	
	P1347R	1347	-6.440	P1229R (-6.538)	No	A (-6.168)	7X (5X)	

[1]Variants are listed using HGVS (Human Genome Variation Society) notation. For example, P1989L indicates P at position 1989 in the moa-tinamou ancestral reference amino acid sequence is replaced by L in moa.
[2]Indicates whether moa amino acid replacement is shared by other birds in alignment.
[3]Indicates alternative amino acid replacement present in other birds in alignment, with PROVEAN score relative to the moa-tinamou ancestor reference in brackets.
[4]Where different, moa raw depth of coverage (DoC) is followed by DoC counting overlapping members of a single read pair as 1X (in brackets).

Suppl. Table S4 Putative function-altering variants in the Galapagos cormorant (*P. harrisi*) are not shared with other flightless lineages

Gene	*P. harrisi* residue	Alignment position (AA)	Amino acid				
			P. harrisi	Moa	Other ratites	Penguins	Flighted birds
DCHS1	2063	2107	D	G	G	G	G
DVL1	103	103	L	P	P	n/a	P
DYNC2H1	2733	2735	S	P	P	P	P
EVC	341	343	I	T	T	n/a	T
FAT1	1717	1742	L	S	S	S	S
	2462	2487	C	Y	Y	Y	Y
GLI2	1086	1127	T	P	P	P/S	P/S
IFT122	691	924	L	Q	Q	Q	Q
KIF7	833	965	W	R	R	R	R
OFD1	325	326	C	R	R	R	R/C
	517	518	T	K	K	K	K/E
	899	924	G	E	E	E	E/N
TALPID3	758	1005	V	D	D	D	D/N
WDR34	188	190	R	P	P	n/a	P