

Congrains, C., Carvalho, A. F., Miranda, E. A., Cumming, G. S., Henry, D. A. W., Manu, S. A., Abalaka, J., Rocha, C. D., Diop, M. S., Sá, J., Monteiro, H., Holbech, L. H., Gbogbo, F. and Del Lama, S. N. 2016. Genetic and paleomodelling evidence of the population expansion of the cattle egret *Bubulcus ibis* in Africa during the climatic oscillations of the Late Pleistocene. – J. Avian Biol. doi: 10.1111/jav.00972

Supplementary material

Appendix 1

Table A1. Location of sampled breeding colonies of cattle egret in Africa.

Locality	Geographical coordinates	
Senegal (SE)		
Zoo - Dakar (ZO)	N 14°43'39.18"	W 17°25'53.70"
Parc Somone (PS)	N 14°29'51.66"	W 17°04'55.86"
Thies (TH)	N 14°47'47.34"	W 16°55'52.98"
Guinea-Bissau (GB)		
Ilha de Bandim (IB)	N 12°07'10.92"	W 15°40'12.84"
Atanque (AT)	N 12°25'47.64"	W 16°05'25.98"
Ilha do Patrão (IP)	N 11°15'49.62"	W 15°46'28.80"
Nigeria (NG)		
Fobour Kasa (FK)	N 09°51'51.62"	E 09°2'44.95"
Kurra Falls (KF)	N 09°25'53.51"	E 09°00'2.27"
Fusa (FU)	N 09°52'47.68"	E 08° 58'15.49"
Toro (TO)	N 10°2'4.99"	E 09°02'19.10"
Ghana (GN)		
Korle Lagoon (KL)	N 05°32'17.51"	W 00°13'12.14"
Bansom Colony (BC)	N 05°06'28.50"	W 01°54'12.60"
South Africa (SA)		
Paarl (PA)	S 33°41'3.06"	E 18°59'1.86"
Boschenmeer Golf State (BG)	S 33°45'00"	E 18°59'00"
Rondevlei (RO)	S 34°03'45.54'	E 18°29'43.5"

Appendix 2

Detection of potential homoplastic sites in CR

The minimum number of recombination events was determined the test described by Hudson and Kaplan (1985) to detect nucleotide sites that may contain homoplastic information. Nucleotide diversity was plotted before and after the removal of suspected homoplastic sites using a sliding window of 20 pb across the CR sequence in the DNAsp 5.1 program (Librado and Rozas 2009) (Fig. S1). Sites with four gametic types in two or more different pairs of sites and occurring in more variable regions were excluded.

The Hudson-Kaplan test revealed 45 sites pairs with four gametic types involving 17 different sites: (9,314); (14,183); (14,219); (14,336); (55,183); (55,231); (122,219); (156,183); (156,219); (156,237); (156,316); (156,319); (156,336); (156,338); (183,219); (183,231); (183,235); (183,257); (183,314); (183,316); (183,319); (183,336); (183,338); (183,346); (219,231); (219,257); (219,314); (219,316); (219,319); (219,336); (231,237); (231,257); (231,314); (231,316); (237,257); (237,316); (257,316); (258,316); (314,316); (314,336); (316,319); (316,338); (319,336); (319,338); (319,346).

Nine nucleotide sites (156, 183, 219, 231, 257, 314, 316, 319 and 336) appeared more frequently in pair sites with four gametes and in the most variable regions. These sites were considered homoplastic and were excluded from of the analysis. After this exclusion, Hudson-Kaplan test did not show site pairs with four gametes.

References

- Hudson, R. R. and Kaplan, N. L. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. - *Genetics*, 111: 147-164.
- Librado, P. and Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. - *Bioinformatics*, 25: 1451-1452.

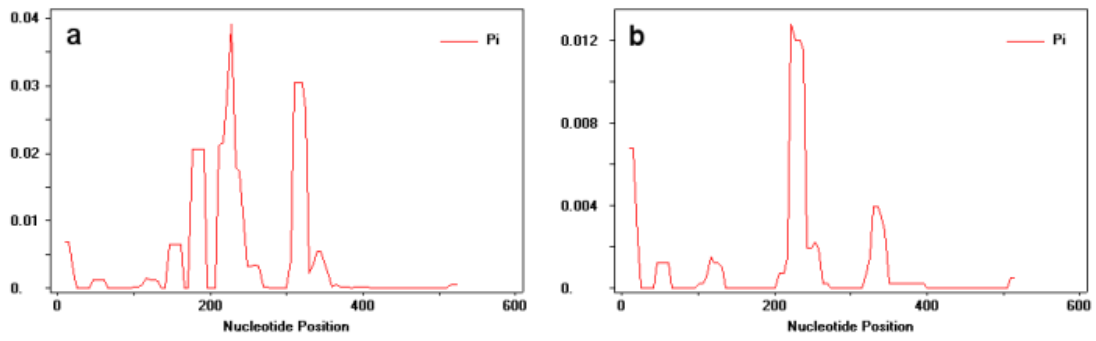


Figure A1. Nucleotide diversity across the CR using 20 pb as the sliding window. (a) Plot before excluding suspected homoplasic sites. (b) Plot before excluding suspected homoplasic sites.

Appendix 3

Haplotypes of CR, ATPase and TGFB2 found in each breeding colony of cattle egret

Table A2. Haplotypes of mtDNA control region found in each breeding colony of cattle egret in Africa. Abbreviations of each sampled location are available in Table S1.

	Senegal			Guinea-Bissau				Nigeria				Ghana		South Africa	
	ZO	PS	TH	IB	AT	IP	FK	KF	FU	TO	KL	BC	PA	BG	RO
H1	20	17	22	20	19	4	20	18	18	20	28	17	20	19	17
H2	5	0	1	2	1	0	0	3	0	0	0	3	0	1	2
H3	0	0	0	1	0	0	1	0	0	0	2	0	0	0	0
H4	7	1	6	1	5	1	0	1	2	1	5	0	3	0	0
H5	3	0	2	0	3	1	0	0	1	0	1	0	0	2	1
H6	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0
H7	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
H8	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0
H9	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H10	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
H11	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
H12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H13	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H14	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
H15	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
H16	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
H17	0	0	0	0	0	0	4	0	2	1	0	0	0	0	0
H18	0	0	0	0	0	0	1	0	0	0	0	0	0	4	2
H19	0	0	0	0	0	0	0	1	1	2	0	0	2	1	2
H20	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
H21	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

Table A3. Haplotypes of ATPase found in each breeding colony of cattle egret in Africa.

Abbreviations of each sampled location are available in Table S1.

	Senegal			Guinea-Bissau				Nigeria			Ghana		South Africa		
	ZO	PS	TH	IB	AT	IP	FK	KF	FU	TO	KL	BC	PA	BG	RO
H1	1	2	0	2	0	1	1	0	0	1	0	1	1	0	0
H2	9	2	4	6	5	2	6	5	2	6	7	6	2	7	7
H3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
H4	0	1	0	0	2	0	0	0	2	0	2	0	0	0	0
H5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
H6	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
H7	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
H8	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
H9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
H10	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
H11	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
H12	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
H13	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Table A4. Haplotypes of TGFB2 found in each breeding colony of cattle egret in Africa.

Abbreviations of each sampled location are available in Table S1.

	Senegal			Guinea-Bissau				Nigeria			Ghana		South Africa		
	ZO	PS	TH	IB	AT	IP	FK	KF	FU	TO	KL	BC	PA	BG	RO
H1	1	1	2	1	1	0	1	0	0	1	0	1	0	1	0
H2	9	4	4	8	7	2	5	3	5	3	9	6	3	2	3
H3	7	3	4	6	5	1	4	4	4	4	2	4	7	7	2
H4	5	0	0	2	1	0	2	1	3	1	1	3	3	0	0
H5	2	0	0	1	1	1	1	0	1	1	2	0	3	1	1
H6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
H7	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
H8	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Appendix 4

Genetic Differentiation analysis

Table A5. Pairwise F_{ST} based on CR among cattle egret populations in colonies of Western and Southern Africa. Significant F_{ST} values after Benjamini-Yekutieli correction are shown. Abbreviations of each sampled location are available in Table S1.

	IB-GB	AT-GB	IP-GB	ZO-SE	PS-SE	TH-SE	KL-GN	BC-GN	FK-NG	KF-NG	FU-NG	TO-NG	PA-SA	BG-SA	RO-AS
IB-GB	-														
AT-GB	0.0172	-													
IP-GB	0.0450	-0.0877	-												
ZO-SE	0.0025	-0.0006	-0.0496	-											
PS-SE	-0.0074	0.0237	0.0565	0.0164	-										
TH-SE	0.0129	-0.0036	-0.0485	-0.0063	0.0142	-									
KL-GN	0.0207	0.0152	-0.0286	0.0176	0.0073	0.0075	-								
BC-GN	-0.0155	0.0306	0.0495	0.0164	0.0214	0.0409	0.0514	-							
FK-NG	0.0331	0.0404	0.0128	0.0534**	0.0182	0.0576**	0.0191	0.0455	-						
KF-NG	-0.0088	0.0178	0.0071	0.0140	-0.0046	0.0335	0.0162	-0.0052	0.0100	-					
FU-NG	0.0000	0.0168	0.0011	0.0109	-0.0155	0.0076	0.0104	0.0244	0.0053	0.0050	-				
TO-NG	0.0149	0.0349	0.0311	0.0240	0.0169	0.0145	0.0141	0.0410*	0.0226	0.0252	-0.0046	-			
PA-SA	0.0296	0.0075	-0.0455	0.0129	0.0247	0.0002	-0.0112	0.0538*	0.0267	0.0194	0.0115	-0.0151	-		
BG-SA	0.0737	0.0611*	-0.0102	0.0646**	0.0654	0.0692*	0.0400	0.0631	0.0338	0.0462	0.0611	0.0480	0.0290	-	
RO-AS	0.0293	0.0381	-0.0246	0.0336	0.0431	0.0388	0.0362	0.0221	0.0345	0.0229	0.0331	0.0013	0.0022	-0.0102	-

*p value > 0.05

**p value > 0.01

Table A6. Pairwise F_{ST} based on ATPase among cattle egret populations in colonies of Western and Southern Africa. There were no significant F_{ST} values after Benjamini-Yekutieli correction ($p < 0.05$). Abbreviations of each sampled location are available in Table S1.

	IB-GB	AT-GB	IP-GB	ZO-SE	PS-SE	TH-SE	KL-GN	BC-GN	FK-NG	KF-NG	FU-NG	TO-NG	PA-SA	BG-SA	RO-SA
IB-GB	-														
AT-GB	0.0788	-													
IP-GB	-0.2233	-0.1048	-												
ZO-SE	-0.0648	0.1161	-0.1818	-											
PS-SE	0.0973	-0.0894	-0.1651	0.1654	-										
TH-SE	0.0315	0.1519	0.0407	0.0241	0.3000	-									
KL-GN	-0.0227	-0.0612	-0.1635	-0.0067	0.0124	0.0542	-								
BC-GN	-0.0539	0.1169	-0.1354	-0.0574	0.2000	-0.0089	-0.0032	-							
FK-NG	-0.0958	0.1033	-0.1667	-0.1036	0.2090	0.0107	-0.0296	-0.1040	-						
KF-NG	-0.0759	0.0663	-0.2000	-0.1340	0.1153	-0.0091	-0.0306	-0.0713	-0.1025	-					
FU-NG	0.1084	-0.0910	-0.0938	0.1505	-0.0895	0.1476	-0.0147	0.1393	0.1250	0.0826	-				
TO-NG	-0.0958	0.1033	-0.1667	-0.1036	0.2090	0.0107	-0.0296	-0.1040	-0.1667	-0.1025	0.1250	-			
PA-SA	-0.0423	0.0444	-0.2063	-0.0205	0.0517	0.0000	-0.0144	-0.1045	-0.0456	-0.0671	0.0504	-0.0456	-		
BG-SA	-0.0539	0.1169	-0.1354	-0.0574	0.2000	-0.0089	-0.0032	-0.0667	-0.1040	-0.0713	0.1393	-0.1040	-0.0291	-	
RO-SA	0.0474	0.2208	0.3000	0.0128	0.4531	0.0729	0.0870	-0.0182	0.0000	0.0278	0.2222	0.0000	0.0729	-0.0182	-

Table A7. Pairwise F_{ST} based on TGFB2 among cattle egret populations in colonies of Western and Southern Africa. There were no significant F_{ST} values after Benjamini-Yekutieli correction ($p < 0.05$). Abbreviations of each sampled location are available in Table S1.

	IB-GB	AT-GB	IP-GB	ZO-SE	PS-SE	TH-SE	KL-GN	BC-GN	FK-NG	KF-NG	FU-NG	TO-NG	PA-SA	BG-SA	RO-SA
IB-GB	-														
AT-GB	-0.0527	-													
IP-GB	-0.1209	-0.1131	-												
ZO-SE	-0.0302	-0.0079	-0.1204	-											
PS-SE	-0.0532	-0.0808	-0.0601	0.0219	-										
TH-SE	-0.0239	-0.0497	-0.0137	0.0362	-0.1041	-									
KL-GN	-0.0150	-0.0225	-0.1542	0.0155	0.0025	0.0701	-								
BC-GN	-0.0614	-0.0511	-0.1068	-0.0434	-0.0439	-0.0212	-0.0051	-							
FK-NG	-0.0420	-0.0236	-0.1310	-0.0541	-0.0001	0.0085	0.0106	-0.0575	-						
KF-NG	-0.0835	-0.0629	-0.1089	-0.0520	-0.0714	-0.0480	0.0016	-0.0886	-0.0667	-					
FU-NG	-0.0228	0.0006	-0.1628	-0.0485	0.0494	0.0761	-0.0166	-0.0329	-0.0550	-0.0296	-				
TO-NG	-0.0574	-0.0419	-0.1348	-0.0591	-0.0266	-0.0349	0.0251	-0.0649	-0.0781	-0.0896	-0.0456	-			
PA-SA	0.0566	0.0957	-0.0655	-0.0049	0.1545	0.1429	0.1228	0.0443	-0.0189	0.0335	-0.0175	-0.0304	-		
BG-SA	0.0108	0.0350	-0.0378	-0.0034	0.0486	0.0169	0.1258	0.0022	-0.0379	-0.0384	0.0237	-0.0740	-0.0125	-	
RO-SA	-0.1039	-0.0992	-0.2477	-0.0757	-0.0813	-0.0312	-0.1099	-0.0898	-0.0855	-0.1112	-0.0944	-0.0971	0.0054	-0.0110	-

Figure A2. SAMOVA calculated for two, three, four and five groups based on mtDNA regions (ATPase and CR) and a nuDNA intron (TGFB2) in African cattle egret populations. X-axis is the genetic differentiation and Y-axis is the number of groups.

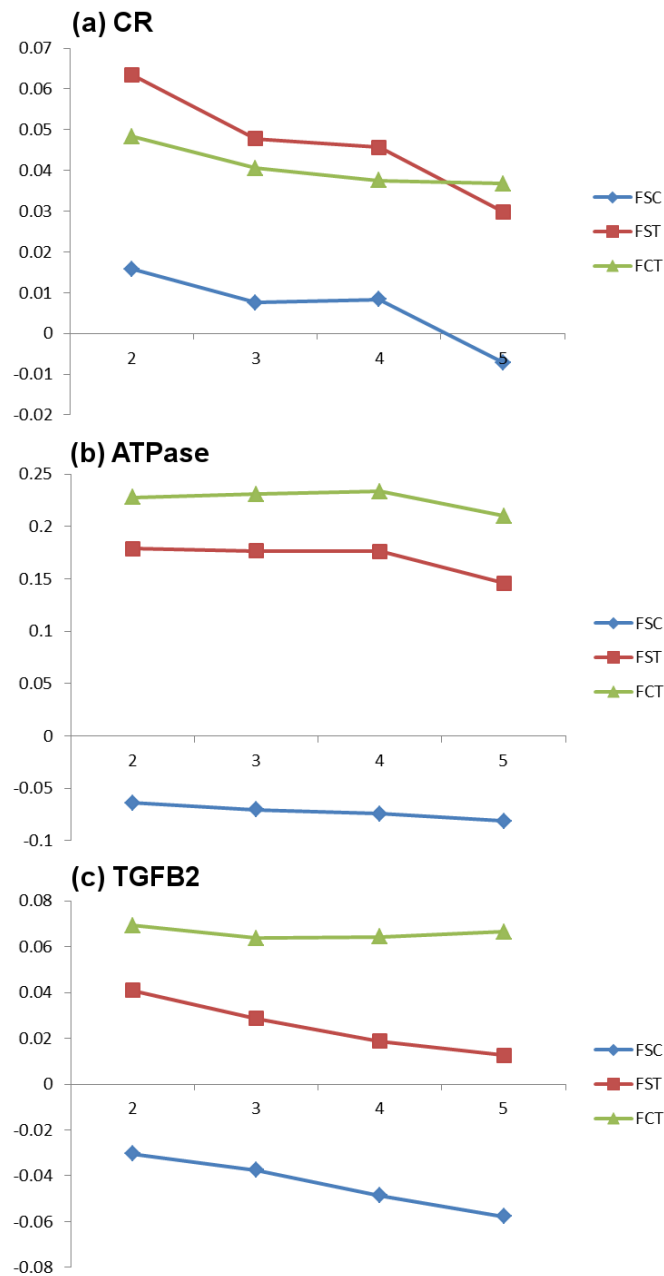
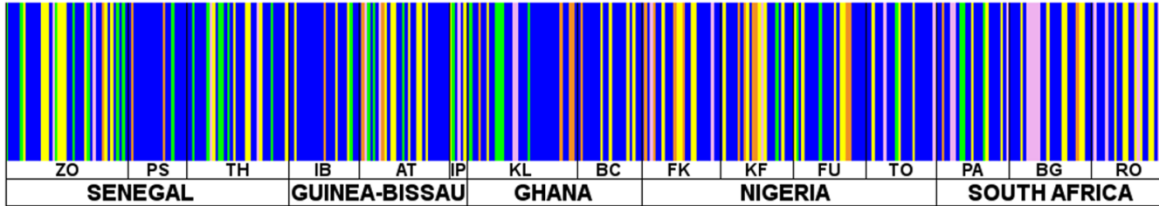


Table A7. SAMOVA calculated for two groups based on mtDNA regions (ATPase and CR) and a nuDNA intron (TGFB2) in African cattle egret populations. Abbreviations of each sampled location are available in Table S1.

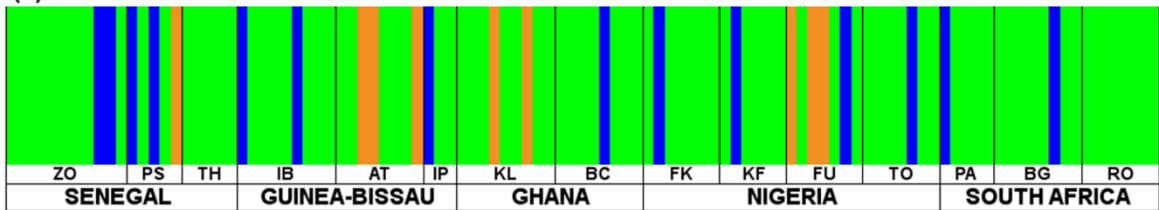
Structure	Percentage of variation	Fixation index	p-value
CR (BG-SA vs other colonies)			
Among groups	4.84	$F_{CT} = 0.04838$	< 0.0001
Among colonies within groups	1.51	$F_{SC} = 0.01584$	0.0020
Within colonies	93.65	$F_{ST} = 0.06346$	< 0.0001
ATPase (AT-GB, PS-SE & FU-NG vs other colonies)			
Among groups	22.79	$F_{CT} = 0.22786$	< 0.0001
Among colonies within groups	-4.93	$F_{SC} = -0.0638$	0.9756
Within colonies	82.14	$F_{ST} = 0.1786$	0.1676
TGFB2 (PA-SA vs other colonies)			
Among groups	6.92	$F_{CT} = 0.06917$	< 0.0001
Among colonies within groups	-2.83	$F_{SC} = -0.03044$	0.9487
Within colonies	95.92	$F_{ST} = 0.04084$	0.8173

Figure A3. Genetic Bayesian clustering based on mtDNA regions (CR and ATPase) and nuDNA (TGFB2) data in cattle egret populations in West and South Africa. Abbreviations of each sampled location are available in Table S1. (a) Five clusters for CR ($p = 1$); (b) three clusters for ATPase ($p = 0.92$); (c) five clusters for TGFB2 ($p = 1$).

(a) CR



(b) ATPase



(c) TGFB2

