Journal of Avian Biology

JAB5113

Mardon, J., Saunders, S. M. and Bonadonna, F. 2011. From preen secretions to plumage: the chemical trajectory of blue petrels' *Halobaena caerulea* social scent. – J. Avian Biol. 42: 29–38.

Supplementary materials: Appendix 1, statistical methods

Chromatographic data from our study were characterised by a large number of variables (i.e. peak areas for all analytes) compared to the number of sample units (n ≤ 72). In addition, the relative abundances of the chemical analytes were rarely normally distributed, typically displaying a high right-skewness. Unfortunately, MANOVA test statistics are not particularly robust to departures from the assumption of multivariate normality (Olson 1974) and simply cannot be computed when there are more variables than sampling units (Anderson 2001). Thus a number of more robust distance-based multivariate approaches which are described below were used instead.

Data pre-treatment, resemblance measure and ordination

Peak areas were successively standardised twice across all samples. The first standardisation used the peak area of the spike (2-bromophenol), to account for variation in the instrument response among samples (particularly across years). The second standardisation used the peak area of a particular target analyte (#265: dodecanoic acid, hexadecyl ester, RI = 3 045), which was one of the highest (if not the highest) peak in all samples. This relativised the values for different analytes within a sample in order to account for the total quantity of secretion, which varied among samples. After standardisation, data were square-root transformed to reduce skewness and so that the resemblance measure calculations, while retaining the relative abundances of analytes, would not be overly dominated by the most abundant analytes (Clarke and Warwick 2001).

Euclidean distances between every pair of samples were then calculated to produce a resemblance matrix that formed the basis of ensuing analyses. Note that Euclidean distance was considered an appropriate choice here, because analytes were measured in similar units and were on similar scales after transformation. In addition, the joint absence of any given analyte was considered to indicate similarity between two samples, and Euclidean distances do not exclude joint absence information. As an illustration, a chemical sexual dimorphism may lie in the systematic absence of certain analytes in one sex compared to the other.

Principal coordinates analysis based on the Euclidean resemblance matrix (PCO`; Gower 1966) was used as an ordination method in order to visualise the patterns of differences in the multivariate chemical structure among samples. Note that although PCO on a Euclidean distance matrix is equivalent to a PCA on the original data, we used PCO here because of the intrinsic over-parameterisation of the problem (many more variables than sampling units).

Statistical methods

We used two different types of distance-based multivariate approaches in our study, PERMANOVA (PERmutational Multivariate ANalysis Of VAriance, Anderson 2001, Mc Ardle and Anderson 2001) and CAP (Canonical Analysis of Principal coordinates, Anderson and Willis 2003). These two types of analysis (PERMANOVA and CAP) offer two alternative statistical perspectives on the data. PERMANOVA indicates how the various factors included in the model contribute to the overall variation in the data. As such, the importance of a given factor is influenced by the quantity of overall variation in the data. CAP models, on the other hand, search the multivariate space for a separation between a priori groups, which can then be used for predictive modelling. This kind of analysis is particularly useful when the direction of segregation between the groups of interest in the multivariate space is fundamentally different from the main direction(s) of overall variation in the dataset (Anderson and Willis 2003) which is the case for the 'Sex' factor in the present study.

CAP models that had a good discriminating capability between groups were used to identify the key analytes associated with the various chemical signals. This was done by examining the linear relationships between each of the individual variables (analytes) and the discriminating axes of the corresponding CAP analysis. In each case, we retained the first 20 analytes which had a Pearson correlation r to the CAP axis higher than a specific threshold value. This specific value was calculated to correspond to the minimum level of correlation that would be deemed statistically significant (after correction for the number of variables tested) in a classical linear correlation analysis (for instance $n_{analytes} = 330$, $n_{samples} = 64$, $r_{min} = 0.45$). This procedure provides correlation-based chemical associations between compounds and the different signals which should not be interpreted in a causative way.

The statistical methods used in each of the three sections of our analysis are described in the main text. The following paragraphs only provide the little extra information which does not appear in the main text.

Chemical trajectory from uropygial secretions to feathers

For the comparison of the chemical profiles from secretion and feather samples, using a single factor PERMANOVA analysis, p-values were obtained using 9 999 permutations of the raw data and Type I (sequential) sum of squares. The analysis was applied to two different datasets: (1) the first one included all the variables (n = 330 analytes), (2) the second one included only those compounds that were found in both sample types and in both years (n = 253 analytes).

The chemical differentiation between sample types was investigated further using a CAP analysis which is a distance-based dis-

criminant analysis, in this case yielding a model to discriminate between sample types on the basis of their chemical profile. Again, this analysis was applied to the two different datasets mentioned above. A leave-one-out cross-validation method was used to determine the number of PCO axes to use for the CAP models (Anderson and Robinson 2003) and to assess their predictive capability.

Presence of sociochemical information on feathers

For the PERMANOVA model used in this section, which included the two factors'Sex' (fixed) and'Individual identity' (random, nested within Sex), p-values were obtained using 9999 permutations of residuals under a reduced model (Freedman and Lane 1983). The design was unbalanced and Type I (sequential) sum of squares were used.

References

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. – Austral. Ecol. 26: 32–46.

Anderson, M. J. and Robinson, J. 2003. Generalized discriminant analysis based on distances. – Australian & New Zealand J. Stat. 43: 75–88.

Anderson, M. J. and Willis, T. J. 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. – Ecology. 84: 511–525.

Clarke, K. R. and Warwick, R. M. 2001. Changes in marine communities: an approach to statistical analysis and interpretation. – Primer-E., Plymouth, UK.

Freedman, D. and Lane, D. 1983. A nonstochastic interpretation of reported significance levels. – J. Bus. and Econ. Stat. 1: 292–298.

Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. – Biometrika 53: 325–338.

McArdle, B. H. and Anderson, M. J. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. – Ecology 82: 290–297.

Olson, C. L. 1974. Comparative robustness of six tests in multivariate analysis of variance. – J. Am. Stat. Ass. 69: 894–908.

Appendix 2

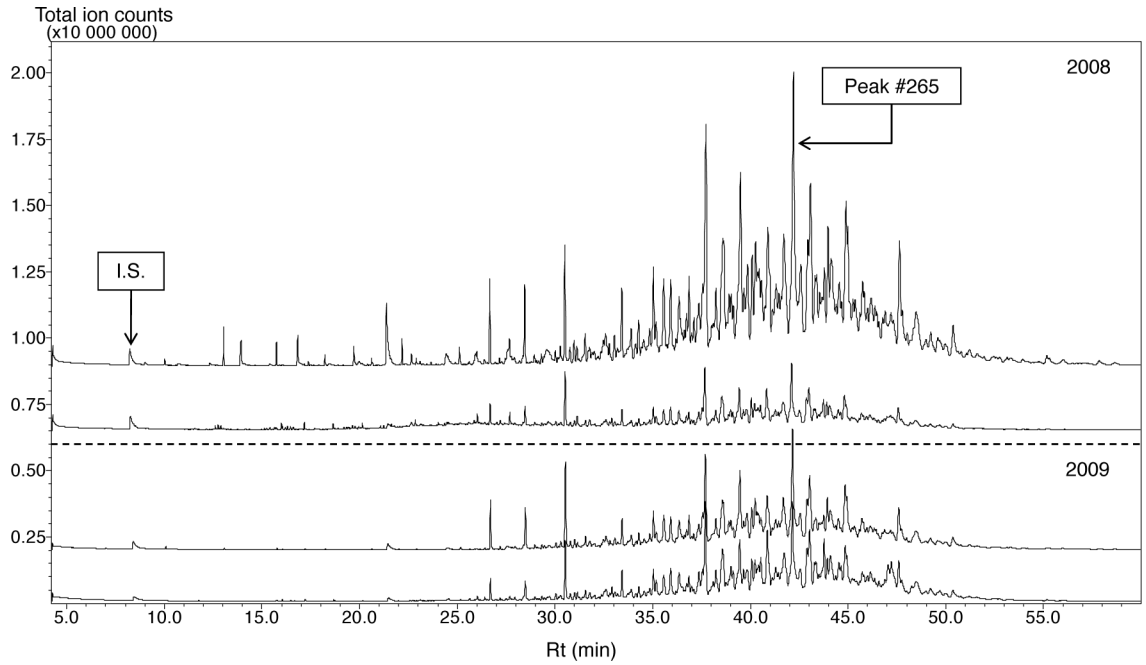

Fig. S1. Full chromatograms obtained with the 4 samples from the same bird (from top to bottom: 2008 secretion, 2008 feather, 2009 secretion, 2009 feather). The two peaks used for standardisation, i.e. the internal standard (I.S.) and the most abundant analyte (peak #265), are indicated.

Appendix 3: tables of chemical composition

Table A3-1: List of all feather-specific analytes, sorted by likely origin

Peak ID#	RI	Best identification
		Secretion-related compounds
10	1315	Nonanoic acid
11	1360	Iso-Decanoic acid
13	1400	n-Tetradecane
14	1405	n-Decanoic acid
19	1500	n-Pentadecane
23	1505	Undecanoic acid, 2-methyl
26	1585	n-Dodecanoic acid
27	1595	1-Tridecanol
37	1705	Hexadecane, 2,6,10-trimethyl
39	1720	Pentadecanal
40	1730	Iso-Tetradecanoic acid, dimethyl ester
43	1775	Tetradecanoic acid
49	1820	Hexadecanal
57	1865	Octadecane, 2-methyl
70	1970	n-Hexadecanoic acid
77	2020	9-Octadecen-1-ol
80	2055	Iso-Hexadecen-1-ol acetate
82	2060	Iso-Heneicosane
84	2065	Eicosane, 2-methyl
85	2070	Iso-Nonadecanol
88	2085	2-Nonadecanol
99	2160	Heneicosane, 5-methyl
103	2175	Iso-Docosane
104	2175	Octadecanoic acid
107	2190	Octadecanamide
109	2205	Iso-Docosane
114	2235	Iso-Tricosane
119	2260	Docosane, 2,21-dimethyl
121	2270	Iso-Heneicosanol
148	2375	Nonadecanamide
155	2400	Tetracosane
		Environmental pollutants
33	1635	Benzophenone
38	1715	Benzoic acid, 2-ethylhexyl ester
54	1850	Benzene, (1-propyldecyl)
58	1875	Benzene, (1-ethylundecyl)
76	2020	Ambreinolide(cis-A/B)
122	2270	Padimate O
		Unresolved origin
56	1860	Unidentified peak
123	2275	Tributyl acetylcitrate
124	2285	Unidentified peak
138	2335	15-Isobutyl-(13αH)-isocopalane
252	2945	Unidentified peak
297	3290	Cholestane-3,5-diol, 5-acetate
333	4360	Iso-Dodecanoic acid, propanetriyl ester

Table A3-2: Main analytes associated with the chemical differentiation between uropygial secretion and feather signals

Peak ID #	RI	Key target analytes: Best identification	Formula	Average peak areas (transformed & standardised): Secretions Mean	Secretions ± SE	Feathers Mean	Feathers ± SE	r_1	r_2
36	1700	n-Heptadecane	$C_{17}H_{36}$	0.8	0.1	114.1	14.1	**0.95**	**0.94**
297	3290	Cholestane-3,5-diol, 5-acetate	$C_{29}H_{50}O_3$	0.0	0.0	107.2	14.9	**0.91**	*NA*
300	3295	Iso-Cholestanol	$C_{27}H_{48}O$	0.3	0.2	247.8	30.5	**0.91**	**0.93**
121	2270	Iso-Heneicosanol	$C_{21}H_{46}O$	0.0	0.0	14.5	2.2	**0.88**	*NA*
305	3350	Unidentified peak	*NA*	14.5	1.3	138.0	15.6	**0.87**	**0.91**
252	2945	Unidentified peak	*NA*	0.0	0.0	29.6	4.1	**0.87**	*NA*
155	2400	Iso-Tetracosane	$C_{24}H_{50}$	0.0	0.0	28.9	3.7	**0.87**	*NA*
19	1500	n-Pentadecane	$C_{15}H_{32}$	0.0	0.0	25.9	3.8	**0.85**	*NA*
37	1710	Iso-Octadecane	$C_{18}H_{38}$	0.0	0.0	4.5	0.7	**0.84**	*NA*
62	1900	n-Nonadecane	$C_{19}H_{40}$	3.0	0.6	62.3	10.5	**0.84**	**0.83**
27	1590	1-Tridecanol	$C_{13}H_{28}O$	0.0	0.0	6.9	1.2	**0.82**	*NA*
85	2070	Iso-Nonadecanol	$C_{19}H_{38}O$	0.0	0.0	13.8	2.4	**0.81**	*NA*
46	1800	n-Octadecane	$C_{18}H_{38}$	1.3	0.2	56.1	9.9	**0.80**	**0.78**
82	2060	Iso-Heneicosane	$C_{21}H_{44}$	0.0	0.0	8.2	1.3	**0.80**	*NA*
72	1990	Iso-Octadecanol	$C_{18}H_{38}O$	0.7	0.1	8.3	1.3	**0.79**	**0.82**
99	2155	Heneicosane, 5-methyl	$C_{22}H_{46}$	0.0	0.0	12.5	2.3	0.79	*NA*
13	1400	n-Tetradecane	$C_{14}H_{30}$	0.0	0.0	8.2	1.2	**0.78**	*NA*
44	1795	Iso-Hexadecanol	$C_{16}H_{34}O$	6.4	1.0	32.8	3.6	**0.78**	**0.82**
48	1805	Benzene, 1-methylundecyl	$C_{18}H_{30}$	0.0	0.0	18.5	3.7	**0.77**	*NA*
29	1600	n-Hexadecane	$C_{16}H_{34}$	1.0	0.1	27.2	4.8	**0.76**	**0.74**
59	1875	Phthalic acid, diisobutyl ester	$C_{16}H_{22}O_4$	3.5	0.5	68.3	13.7	**0.75**	**0.73**
35	1680	Iso-Hexadecanol	$C_{16}H_{34}O$	2.5	0.5	17.9	2.6	**0.75**	**0.76**
52	1830	Benzene, (1-pentylheptyl)	$C_{18}H_{30}$	0.0	0.0	13.3	3.1	**0.75**	*NA*
101	2175	n-Pentadecylcyclohexane	$C_{21}H_{42}$	0.4	0.1	24.8	5.4	**0.71**	**0.70**
42	1755	Iso-Hexadecanol	$C_{16}H_{34}O$	1.1	0.3	16.4	3.5	**0.66**	**0.71**

Note: *r* corresponds to the Pearson correlation of a particular compound with the CAP axis discriminating the two sample types (r_1 is from the first CAP model including all analytes, r_2 is from the second model limited to analytes common to both sample types). All contributions presented are significant (critical *r* value, at a level of α=5%, was 0.45).

Table A3-3: Main analytes associated with the chemical 'Sex' badge in secretions, feathers and all samples together

Key analytes					Pearson *r* (with CAP axis)			
Peak ID #	RI	Best Identification	Methyl subt.	Formula	Secretions	Feathers	All samples	Signal direction
247	2920	Iso-Decanoic acid, octadecyl ester	4	$C_{28}H_{56}0_2$	**0.93**	**0.73**	**0.79**	Females
201	2650	Iso-Nonanoic acid, hexadecyl ester	2-4	$C_{25}H_{50}O_2$	**0.89**	**0.68**	**0.78**	Females
246	2910	Iso-Undecanoic acid, heptadecyl ester	2-4	$C_{28}H_{56}O_2$	**0.89**	**0.64**	**0.72**	Females
239	2870	Iso-Undecanoic acid, hexadecyl ester	2-4	$C_{27}H_{54}O_2$	**0.87**	**0.65**	**0.70**	Females
208	2685	Iso-Decanoic acid, pentadecyl ester	4	$C_{25}H_{50}O_2$	**0.87**	0.53	**0.70**	Females
230	2820	Iso-Decanoic acid, heptadecyl ester	4	$C_{27}H_{54}O_2$	**0.88**	0.57	**0.67**	Females
199	2645	Iso-Decanoic acid, pentadecyl ester	2-4	$C_{25}H_{50}O_2$	**0.76**	0.58	**0.66**	Females
261	3005	Iso-Undecanoic acid, octadecyl ester	4	$C_{29}H_{58}0_2$	**0.73**	**0.64**	**0.65**	Females
192	2600	Iso-Decanoic acid, tetradecyl ester	4	$C_{24}H_{48}O_2$	**0.83**	**0.64**	**0.64**	Females
223	2780	Iso-Decanoic acid, hexadecyl ester	2-4	$C_{26}H_{52}O_2$	**0.80**	0.54	**0.63**	Females
177	2525	Iso-Nonanoic acid, pentadecyl ester	2-4	$C_{24}H_{48}O_2$	**0.76**	0.52	**0.63**	Females
222	2770	Iso-Undecanoic, pentadecyl ester	2-4	$C_{26}H_{52}O_2$	**0.67**	0.56	**0.62**	Females
253	2955	Iso-Undecanoic, heptadecyl ester	4	$C_{28}H_{56}O_2$	**0.72**	0.60	**0.62**	Females
234	2840	Iso-Hexacosanol		$C_{26}H_{54}O$	0.60	0.49	**0.54**	Females
236	2855	Iso-Nonanoic acid, octadecyl ester	3	$C_{27}H_{54}O_2$	0.55	0.51	**0.53**	Females
160	2440	Iso-Decanoic acid, tridecyl ester	4	$C_{23}H_{46}O_2$	**0.68**	0.49	**0.53**	Females
186	2555	Iso-Decanoic acid, tetradecyl ester	4	$C_{24}H_{48}O_2$	**0.69**	0.47	**0.52**	Females
179	2535	Iso-Octanoic acid, hexadecyl ester	4	$C_{24}H_{48}O_2$	**0.75**	0.41	**0.50**	Females
213	2710	Iso-Dodecanoic acid, tetradecyl ester	2	$C_{26}H_{52}O_2$	–0.55	–0.61	**–0.60**	Males
204	2660	Iso-Dodecanoic acid, tridecyl ester	NB	$C_{25}H_{50}O_2$	–0.49	–0.56	**–0.51**	Males
250	2940	Iso-Dodecanoic acid, hexadecyl ester	NB	$C_{28}H_{56}O_2$	–0.47	–0.53	**–0.50**	Males

Note: *r* corresponds to the Pearson correlation of a particular compound with the CAP axis discriminating the two sexes in the corresponding model. Strong contributions are bold. For information, critical *r* values (at a level of α = 5%) would be respectively 0.62 (secretions), 0.62 (feathers) and 0.45 (all samples).