

Avian *Clock* gene polymorphism: evidence for a latitudinal cline in allele frequencies

A. JOHNSEN,*††† A. E. FIDLER,*‡†† S. KUHN,* K. L. CARTER,* A. HOFFMANN,* I. R. BARR,§¶ C. BIARD,** A. CHARMANTIER,†† M. EENS,‡‡ P. KORSTEN,§§ H. SIITARI,¶¶ J. TOMIUK*** and B. KEMPENAERS*

*Max Planck Institute for Ornithology, Department of Behavioural Ecology & Evolutionary Genetics, D-82305 Seewiesen, Germany, †Natural History Museum, University of Oslo, N-0318 Oslo, Norway, ‡Cawthron Institute, Nelson 7042, New Zealand, §University of East Anglia, Norwich NR4 7TJ, UK, ¶Edward Grey Institute of Field Ornithology, University of Oxford, Oxford OX1 3PS, United Kingdom, **Laboratoire de Parasitologie Evolutive, Université Pierre et Marie Curie, 75252 Paris, cedex 05, France, ††Centre d'Ecologie Fonctionnelle Evolutive, CNRS (UMR 5175), 34293 Montpellier cedex 5, France, ‡‡Department of Biology, University of Antwerp, B-2020 Antwerp, Belgium, §§Animal Ecology Group, University of Groningen, 9750 AA Haren, The Netherlands, ¶¶Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland, ***Institute of Human Genetics, University of Tübingen, FIN-40014 Tübingen, Germany

Abstract

In comparison with most animal behaviours, circadian rhythms have a well-characterized molecular genetic basis. Detailed studies of circadian clock genes in 'model' organisms provide a foundation for interpreting the functional and evolutionary significance of polymorphic circadian clock genes found within free-living animal populations. Here, we describe allelic variation in a region of the avian *Clock* orthologue which encodes a functionally significant polyglutamine repeat (*ClkpolyQc*ds), within free-living populations of two passerine birds, the migratory bluethroat (*Luscinia svecica*) and the predominantly nonmigratory blue tit (*Cyanistes caeruleus*). Multiple *ClkpolyQc*ds alleles were found within populations of both species (bluethroat: 12 populations, 7 alleles; blue tit: 14 populations, 9 alleles). Some populations of both species were differentiated at the *ClkpolyQc*ds locus as measured by F_{ST} and R_{ST} values. Among the blue tit, but not bluethroat populations, we found evidence of latitudinal clines in (i) mean *ClkpolyQc*ds repeat length, and (ii) the proportions of three *ClkpolyQc*ds genotype groupings. Parallel analyses of microsatellite allele frequencies, which are considered to reflect selectively neutral processes, indicate that interpopulation allele frequency variation at the *ClkpolyQc*ds and microsatellite loci does not reflect the same underlying demographic processes. The possibility that the observed interpopulation *ClkpolyQc*ds allele frequency variation is, at least in part, maintained by selection for microevolutionary adaptation to photoperiodic parameters correlated with latitude warrants further study.

Keywords: circadian, clock, *Cyanistes*, *Luscinia*, polyglutamine, polymorphism

Received 2 March 2007; revision received 2 July 2007; accepted 15 August 2007

Introduction

The molecular characterization of an ever-increasing number of behaviour-related genes has made evolutionary studies of such genes in free-living, nonmodel organisms increasingly

feasible (Boake *et al.* 2002; Fitzpatrick *et al.* 2005). Following such a comparative genomic approach, we studied spatial patterns of genetic population differentiation in avian orthologues of a gene central to vertebrate circadian rhythms.

The circadian clock, which generates endogenous behavioural and physiological rhythms, and which also aids in the interpretation of seasonal changes in photoperiod length, is perhaps the aspect of animal behaviour most fully characterized at the molecular genetic level. Intensive

Correspondence: Bart Kempenaers, Fax: 0049 8157932 400; E-mail: b.kempenaers@orn.mpg.de

A. Johnsen and A.E. Fidler contributed equally to this work.

studies of phylogenetically diverse, laboratory-living, model organisms (predominantly fruit flies, *Drosophila melanogaster*, fungi *Neurospora crassa* and mice *Mus musculus*) have produced very similar models of gene networks that interact to generate biochemical oscillations with an intrinsic period of approximately 24 h (Young & Kay 2001; Panda *et al.* 2002; Reppert & Weaver 2002; Ko & Takahashi 2006). Furthermore, evolutionary and functional studies of allelic variants of a *D. melanogaster* circadian clock gene, *period*, provided an early example of the study of behaviour-related genes in free-living animal populations. More specifically, frequencies of *Drosophila period* alleles, varying in the length of a repetitive coding sequence, display north–south (i.e. latitudinal) clines across Europe (reviewed in Tauber & Kyriacou 2005). It has been proposed that such clines are maintained by selection for local adaptation of the *Drosophila* circadian clock to the differing ambient temperature ranges found at differing latitudes (Tauber & Kyriacou 2005).

The vertebrate circadian gene, *Clock*, encodes a protein (CLOCK) which heterodimerizes with a second protein, BMAL1, to produce a transcription-activating complex which forms a core component of the vertebrate circadian oscillator (Young & Kay 2001; Panda *et al.* 2002; Ko & Takahashi 2006). In addition to its role within the core oscillator, CLOCK/BMAL1 mediates oscillator output through transcriptional activation of downstream, clock-controlled genes (Reppert & Weaver 2002; Iuvone *et al.* 2005). Human population genetic studies have reported an association between a *Clock* gene 3'-UTR (untranslated region) single nucleotide polymorphism and variation in sleeping time preferences (Katzenberg *et al.* 1998; Mishima *et al.* 2005). Furthermore, there is evidence that *Clock* gene polymorphisms are associated with interstrain differences in trout spawning times (Leder *et al.* 2006). The CLOCK protein's C-terminal domain is glutamine rich and includes a

polyglutamine (poly Q) repeat, a characteristic feature of many transcription factors (Mitchell & Tjian 1989; Young & Kay 2001; Panda *et al.* 2002). Both behavioural and molecular characterizations of mouse and *Drosophila Clock* gene mutants point to the functional importance of the CLOCK C-terminal region (Young & Kay 2001; Panda *et al.* 2002). Furthermore, *in vitro* studies indicate that the poly Q repeat region influences the transcription activating potential of the CLOCK/BMAL1 heterodimer (Avivi *et al.* 2001; Young & Kay 2001).

In view of the possible biochemical and behavioural consequences of CLOCK protein poly Q repeat length variation, we (i) looked for allelic length variation in the *Clock* poly Q coding sequence (*ClkpolyQc*) within free-living populations of two widespread passerine birds: the migratory bluethroat (*Luscinia svecica*) and the largely nonmigratory blue tit (*Cyanistes caeruleus*); (ii) tested for genetic differentiation of the passerine populations at the *Clock* locus; (iii) examined whether the distribution of the *Clock* allelic variants was related to breeding latitude; and (iv) compared patterns of *Clock* allele frequency variation with analogous variation at presumably selectively neutral microsatellite loci.

Materials and methods

Collection of blood samples and DNA purification

Blood samples were collected from putatively unrelated individuals in 14 European blue tit populations ($n = 1461$ individuals) and 12 European and west Asian bluethroat populations ($n = 369$ individuals) (Fig. 1, Table 1). All birds were sampled at the breeding grounds. Blood samples (10–50 μ L) collected from the brachial vein were suspended in either Queen's lysis buffer [10 mM Tris, 10 mM NaCl, 10 mM EDTA, 1.0% (w/v) *n*-lauroylsarcosine, pH 8.0] or

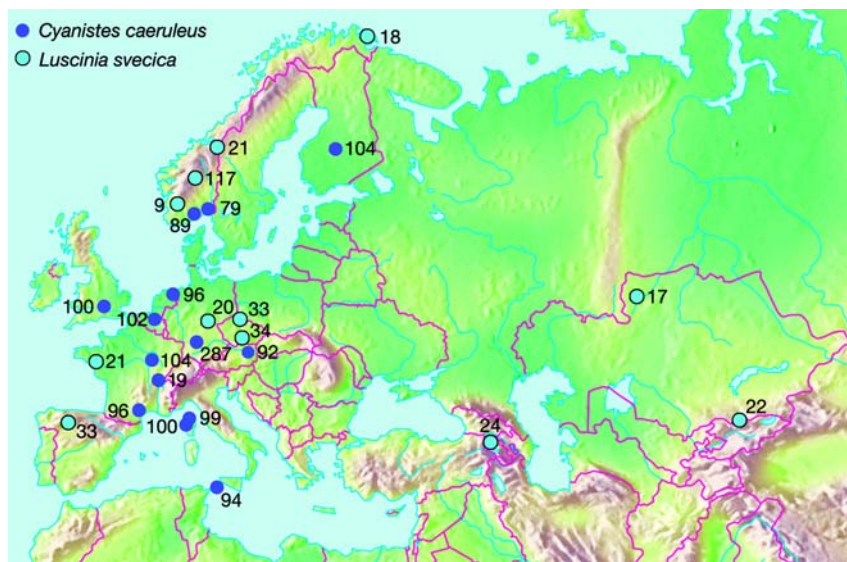


Table 1 Population details, *Clk*polyQc_{ds} allele frequencies and observed heterozygosity for the 12 bluethroat (*Luscinia svecica*) (A) and 14 blue tit (*Cyanistes caeruleus*) (B) populations used in this study

Population	Subspecies	<i>n</i>	Latitude	Longitude	Q ₁₀	Q ₁₁	Q ₁₂	Q ₁₃	Q ₁₄	Q ₁₅	Q ₁₆	<i>H</i>
Sandfjorden, Norway	<i>svecica</i>	18	70°30'N	30°32'E	0	0.028 (1/36)	0	0.889 (32/36)	0.083 (3/36)	0	0	0.222
Stugudal, Norway	<i>svecica</i>	21	62°59'N	12°30'E	0	0.024 (1/42)	0	0.976 (41/42)	0	0	0	0.048
Heimdalen, Norway	<i>svecica</i>	117	61°25'N	8°52'E	0.004 (1/234)	0.017 (4/234)	0.009 (2/234)	0.919 (215/234)	0.047 (11/234)	0	0.004 (1/234)	0.154
Sirdal, Norway	<i>svecica</i>	9	59°09'N	6°55'E	0	0	0	0.889 (16/18)	0.111 (2/18)	0	0	0.222
Kostanay, Kazakhstan	<i>pallidogularis</i>	17	51°38'N	64°13'E	0	0	0	0.853 (29/34)	0.118 (4/34)	0.029 (1/34)	0	0.235
Krkonoše Mts., Czech Rep.	<i>svecica</i>	33	50°50'N	15°35'E	0	0.030 (2/66)	0	0.939 (62/66)	0.030 (2/66)	0	0	0.121
Thüringen, Germany	<i>cyaneola</i>	20	50°21'N	10°44'E	0	0	0	0.725 (29/40)	0.250 (10/40)	0.025 (1/40)	0	0.400
Třeboň, Czech R.	<i>cyaneola</i>	34	48°59'N	14°48'E	0	0	0	0.897 (61/68)	0.103 (7/68)	0	0	0.206
Guérande, France	<i>namnetum</i>	21	47°20'N	2°25'W	0	0	0	0.714 (30/42)	0.286 (12/42)	0	0	0.476
Valduerna, Spain	<i>azuricollis</i>	33	42°20'N	5°58'W	0	0	0	1.000 (66/66)	0	0	0	0.000
Suusamy, Kirgistan	<i>tianshanica</i>	22	42°10'N	73°45'E	0	0.455 (2/44)	0	0.864 (38/44)	0.091 (4/44)	0	0	0.182
Mount Aragats, Armenia	<i>magna</i>	24	40°30'N	44°15'E	0	0	0	0.854 (41/48)	0.146 (7/48)	0	0	0.292

n, number of individuals sampled; *H*, observed heterozygosity.

B. Blue tit (*Cyanistes caeruleus*)

Population	Subspecies	<i>n</i>	Latitude	Longitude	Q ₉	Q ₁₀	Q ₁₁	Q ₁₂	Q ₁₃	Q ₁₄	Q ₁₅	Q ₁₆	Q ₁₇	<i>H</i>
Jyväskylä, Finland	<i>caeruleus</i>	104	62°37'N	26°21'E	0.005 (1/194)	0.005 (1/194)	0.026 (5/194)	0.567 (110/194)	0.351 (68/194)	0.046 (9/194)	0	0	0	0.567
Oslo, Norway	<i>caeruleus</i>	79	59°58'N	10°47'E	0	0	0.006 (1/154)	0.643 (99/154)	0.260 (40/154)	0.091 (14/154)	0	0	0	0.455
Jomfruland, Norway	<i>caeruleus</i>	89	58°52'N	9°36'E	0	0.006 (1/170)	0.076 (13/170)	0.441 (75/170)	0.188 (32/170)	0.288 (49/170)	0	0	0	0.576
Vosbergen, Netherlands	<i>caeruleus</i>	96	53°08'N	6°35'E	0	0.010 (2/192)	0.042 (8/192)	0.641 (123/192)	0.193 (37/192)	0.115 (22/192)	0	0	0	0.479
Wytham Woods, UK	<i>obscurus</i>	100	51°45'N	1°20'W	0	0.010 (2/192)	0.057 (22/192)	0.589 (113/192)	0.307 (59/192)	0.036 (7/192)	0	0	0	0.521
Antwerp, Belgium	<i>caeruleus</i>	102	51°09'N	4°24'E	0	0	0.054 (11/204)	0.598 (122/204)	0.275 (56/204)	0.069 (14/204)	0.005(1/204)	0	0	0.510
Tübingen, Germany	<i>caeruleus</i>	287	48°33'N	9°00'E	0.002 (1/570)	0.005 (3/570)	0.033 (19/570)	0.660 (376/570)	0.251 (143/570)	0.047 (27/570)	0.002 (1/570)	0	0	0.470
Forêt d'Orient, France	<i>caeruleus</i>	104	48°17'N	4°18'E	0.005 (1/206)	0	0.034 (7/206)	0.578 (119/206)	0.272 (56/206)	0.112 (23/206)	0	0	0	0.583
Vienna, Austria	<i>caeruleus</i>	92	48°13'N	16°20'E	0	0.005 (1/182)	0.044 (8/182)	0.522 (95/182)	0.368 (67/182)	0.060 (11/182)	0	0	0	0.637
Auxonne, France	<i>caeruleus</i>	19	47°2'N	5°35'E	0	0	0	0.722 (26/36)	0.167 (6/36)	0.111 (4/36)	0	0	0	0.444
Rouviere, France	<i>caeruleus</i>	96	43°40'N	3°40'E	0	0.016 (3/182)	0.027 (5/182)	0.604 (110/182)	0.313 (57/182)	0.033 (6/182)	0	0.005 (1/182)	0	0.560
Muro, Corsica, France	<i>ogliastreae</i>	99	42°32'N	8°54'E	0	0.021 (4/190)	0.011 (2/190)	0.637 (121/190)	0.189 (36/190)	0.137 (26/190)	0	0	0.005 (1/190)	0.516
Pirio, Corsica, France	<i>ogliastreae</i>	100	42°23'N	8°45'E	0	0.036 (7/192)	0.042 (8/192)	0.672 (129/192)	0.167 (32/192)	0.083 (16/192)	0	0	0	0.531
Pantelleria, Italy	<i>ultramarinus</i>	94	36°44'N	12°00'E	0	0	0	1.000 (180/180)	0	0	0	0	0	0.000

ethanol and stored at room temperature. Genomic DNA was purified using commercially available kits [QIAamp DNA blood kit (QIAGEN); GFX Genomic Blood DNA Purification kit (Amersham Pharmacia Biotech)] and stored at -20°C .

Sequencing of the ClkpolyQcfs alleles

Nucleotide sequences of human (GenBank Accession nos: AF011568, AF097457), mouse (AF000998), rat (AB019258), chicken (AF132531) and quail (AB029889) *Clock* orthologues were aligned and polymerase chain reaction (PCR) primers designed to anneal to conserved sequences located at the predicted 5' and 3' boundaries of a predicted exon, corresponding to human *Clock* gene exon 20 (Steeves *et al.* 1999), encoding the passeriform *ClkpolyQcfs*. Forward primer (0.8 μM): 5'-TTTTCTCAAGGTCAGCAGCTTGT-3' and reverse primer (0.8 μM): 5'-CTGTAGGAAGTGTG(C/T)GG(G/T)TGCTG-3' were used with reaction conditions: 0.2 mM dNTPs, $\text{Mg}^{2+} = 1.5$ mM; $94^{\circ}\text{C}/2$ min; $94^{\circ}\text{C}/30$ s, $58^{\circ}\text{C}/30$ s, $72^{\circ}\text{C}/60$ s, 10 cycles; $94^{\circ}\text{C}/30$ s, $62^{\circ}\text{C}/30$ s, $72^{\circ}\text{C}/60$ s increasing 5 s/cycle, 30 cycles; $72^{\circ}\text{C}/7$ min, hold at 4°C , catalyzed using *Taq* DNA polymerase (Roche Diagnostics) following the manufacturer's instructions. Amplification products were ligated into pGEM-T Easy (Promega) and sequenced by an external contractor (MWG Biotech). Sequence identities were determined by comparison with the chicken and quail *Clock* sequences (BLAST2, www.ncbi.nlm.nih.gov/blast/) and sequence alignments accomplished using CLUSTAL W (<http://align.genome.jp/>).

Screening for avian ClkpolyQcfs alleles

Genomic DNA samples were screened for length polymorphism in the *ClkpolyQcfs* by PCR amplification followed by resolution and detection on a conventional DNA sequencing machine. We used PCR forward primer (5'-labelled with the 'blue' fluorescent dye 6-FAM) 5'-6-FAM-TGGAGCGGTAATGGTACCAAGTA-3' and reverse primer 5'-TCAGCTGTGACTGAGCTGGCT-3'. Amplification reaction conditions for the *Taq* DNA polymerase catalyzed PCR were different for the two species; bluethroat: $\text{Mg}^{2+} = 1.5$ mM; $94^{\circ}\text{C}/5$ min; $94^{\circ}\text{C}/30$ s, $72^{\circ}\text{C}/30$ s, $60^{\circ}\text{C}/30$ s decreasing 2°C each cycle (touchdown PCR), 4 cycles; $94^{\circ}\text{C}/30$ s, $50^{\circ}\text{C}/30$ s, $72^{\circ}\text{C}/30$ s, 25 cycles, hold at 4°C ; blue tit: $\text{Mg}^{2+} = 2.0$ mM; $92^{\circ}\text{C}/2$ min; $92^{\circ}\text{C}/30$ s, $53^{\circ}\text{C}/30$ s, $72^{\circ}\text{C}/30$ s, 25 cycles, $72^{\circ}\text{C}/30$ s, 4°C hold. One or both of two types of molecular weight standards were used for the estimation of amplified allele sizes. The first was a conventional molecular size standard (GeneScan-500 LIZ, Applied Biosystems). The second molecular weight standard was generated 'in-house' from a template consisting of a pool of seven, sequenced plasmids encoding blue tit *ClkpolyQcfs* alleles Q_9 – Q_{15} . Primers were identical to those

used for the genomic samples except that the forward primer was 5'-labelled with the fluorescent dye VIC. Reaction conditions were as described for the blue tit genomic DNA samples. Amplification products from the Q_9 – Q_{15} allele plasmid pool were diluted 1/20 and mixed with PCR products from the genomic DNA being genotyped. Because the fluorescent dyes VIC and 6-FAM emit light of differing wavelengths ('green' and 'blue', respectively), it is possible to distinguish the molecular weight standard molecules (green) from those amplified from genomic samples (blue). Amplification products (about 290 bp) were resolved using an ABI PRISM Genetic Analyzer (Applied Biosystems) and sizes of amplification products calculated using commercial software (GENESCAN 3.7 and GENOTYPER 3.6, Applied Biosystems). We succeeded in genotyping 97% (1422/1461) of the blue tit samples and 100% of the bluethroat samples.

Microsatellite analyses

In order to test whether *ClkpolyQcfs* alleles behaved differently from neutral expectations, we genotyped the blue tit samples at nine polymorphic microsatellite loci [Mcy μ 4, Double *et al.* 1997; Pat MP 2–43, Otter *et al.* 1998; PC3, PC7, PC8, PC9, Dawson *et al.* 2000; PK12, S.M. Tanner, H. Richner and D. Schuenperli, unpublished data; EMBL accession no. AF041466]; Pocc1, Pocc6, Bensch *et al.* 1996], using previously reported PCR conditions and an ABI 3100 Automatic sequencer (see Delhey *et al.* 2003 for details). The nine loci showed moderate to high levels of polymorphism (mean number of alleles $\pm$ SD: 33.1 ± 18.5 , range 15–76). The Pocc6 marker did not amplify a product for individuals from the Pantelleria population. A total of 1342 samples were successfully genotyped at four or more loci (mean $\pm$ SD: 8.6 ± 0.1 loci). For unknown reasons, 39 of these samples did not return a *Clock* genotype. The proportion of the *Clock*-genotyped individuals ($n = 1422$) that were typed for at least four microsatellites ranged from 0.49 to 1.0 in the 14 populations (mean $\pm$ SD = 0.91 ± 0.13). The low success rate of microsatellite typing of the Pantelleria population (49%) did not influence any of the main results of this study, since (i) the population was monomorphic with respect to *Clock* alleles (see below), hence the allele frequency and average allele size is the same irrespective of the specific samples used, and (ii) for reasons explained below, we present results both with and without this particular population. Excluding the Pantelleria population, the average proportion of *Clock*-genotyped individuals typed for at least four microsatellites was 0.95 ± 0.05 (range 0.87–1.0). For the bluethroat, we compared variation in *ClkpolyQcfs* allele frequencies with variation at 11 microsatellite loci, as reported in a recent study of bluethroat phylogeography that used the same samples as this study (Johnsen *et al.* 2006).

Table 2 Mantel tests of spatial patterns in *ClkpolyQc*ds and microsatellite allele length and the frequency of the two most common alleles, using data from 13 blue tit populations (excluding Pantelleria, see text). MCA, most common allele; 2nd MCA, second most common allele. *P* values < 0.05 are shown in bold. Data only shown for the three microsatellites showing signs of latitudinal clines (PC3, PC7 and PK12; see text)

	Geographical distance		Latitudinal distance		Longitudinal distance	
	<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>
<i>ClkpolyQc</i>ds						
Mean allele length	0.24	0.10	0.36	0.018	0.11	0.17
Q ₁₂ (MCA) frequency	-0.23	0.068	-0.28	0.026	-0.18	0.14
Q ₁₃ (2nd MCA) frequency	0.17	0.17	0.11	0.23	0.30	0.056
Microsatellites						
Mean allele length (9 markers)	-0.06	0.35	-0.20	0.077	0.61	0.0003
PC3 mean allele length	-0.13	0.22	-0.28	0.051	0.21	0.11
PC3 MCA frequency	0.01	0.47	-0.04	0.40	0.13	0.23
PC3 2nd MCA frequency	0.08	0.32	0.04	0.40	-0.10	0.25
PC7 mean allele length	0.48	0.006	0.52	0.004	0.34	0.079
PC7 MCA frequency	-0.31	0.029	-0.38	0.011	-0.06	0.28
PC7 2nd MCA frequency	0.36	0.031	0.36	0.024	0.14	0.20
PK12 mean allele length	-0.46	0.004	-0.53	0.001	-0.19	0.13
PK12 MCA frequency	0.22	0.15	0.34	0.030	-0.21	0.072
PK12 2nd MCA frequency	-0.09	0.25	-0.14	0.17	-0.02	0.39

Statistical analyses

For both the blue tit and bluethroat, several recognized subspecies were represented in the samples (see Tables 1 and 2). We treated each population as a separate unit of analysis unless otherwise stated. Based on morphology, the Pantelleria blue tit population is more closely related to North African blue tits than to European ones (Cramp & Perrins 1993), and it has been proposed by some authors that North African blue tits should be treated as a separate species (Salzburger *et al.* 2002; Kvist *et al.* 2005). Therefore, we report two sets of analyses, one including the Pantelleria blue tit population and another excluding it. Genotype frequency data were tested statistically for deviation from Hardy–Weinberg proportions and linkage equilibrium, using GENEPOP 3.4 (web version; <http://genepop.curtin.edu.au/>, with the following Markov chain parameters: dememorization, 10 000; batches, 10 000; iterations per batch, 10 000). Overall genetic differentiation between populations was estimated in two ways. First, we calculated F_{ST} values (Weir & Cockerham 1984) using FSTAT (Goudet 1995), which assumes the infinite allele model of mutation (Kimura & Crow 1964). Second, we calculated the R_{ST} fixation index (Slatkin 1995), using RSTCALC (Goodman 1997), which assumes the stepwise mutation model for polymorphic repeat sequences (Ohta & Kimura 1973). Note that the R_{ST} fixation calculations were only possible for the *ClkpolyQc*ds data since several of the microsatellite loci did not conform to a strict stepwise mutation model. Statistical tests of overall population differentiation were performed for both the F_{ST} and R_{ST} indices, using the

permutation procedures within each corresponding program. Since both fixation indices produced qualitatively similar results (see Table S1, Supplementary material), we only report the results from the F_{ST} analyses. As a rule of thumb, the sample size of each population should be at least twice the number of alleles at each locus, in order to obtain unbiased F_{ST} estimates (S. Bensch, personal communication). In the blue tit microsatellite analysis, there was one population with a low sample size (Auxonne: 16 individuals typed for four or more microsatellite loci) and two loci with high numbers of alleles (PC3, 46 alleles; PC8, 76 alleles). To test whether this influenced our analyses, we repeated the F_{ST} analyses on a restricted data set excluding the Auxonne population and data on PC3 and PC8 for all populations. The Pantelleria population was also excluded from these analyses to avoid the strong influence of this genetically distinct population (see Results). Since (i) the overall level of population differentiation was virtually identical when based on this restricted data set ($F_{ST} = 0.019$) as compared to the full data set ($F_{ST} = 0.018$), and (ii) pairwise F_{ST} estimates based on the restricted data set correlated strongly with those calculated from two random halves of the microsatellite loci (subset 1: Pocc1, Pocc6, PC3, PC7, PC8; subset 2: PC9, Mcyu4, Pat43, PK12) (Mantel tests, both $r > 0.97$, both $P < 0.001$), we only report the analyses based on the full data set.

To test for patterns of isolation by distance, we performed Mantel tests on matrices of pairwise genetic differentiation (i.e. F_{ST} values) and geographical distance. Mantel tests were performed using the software MANTEL 2.0 (<http://www.terc.csiro.au/profile.asp?ID=LIEDA>), with 3000

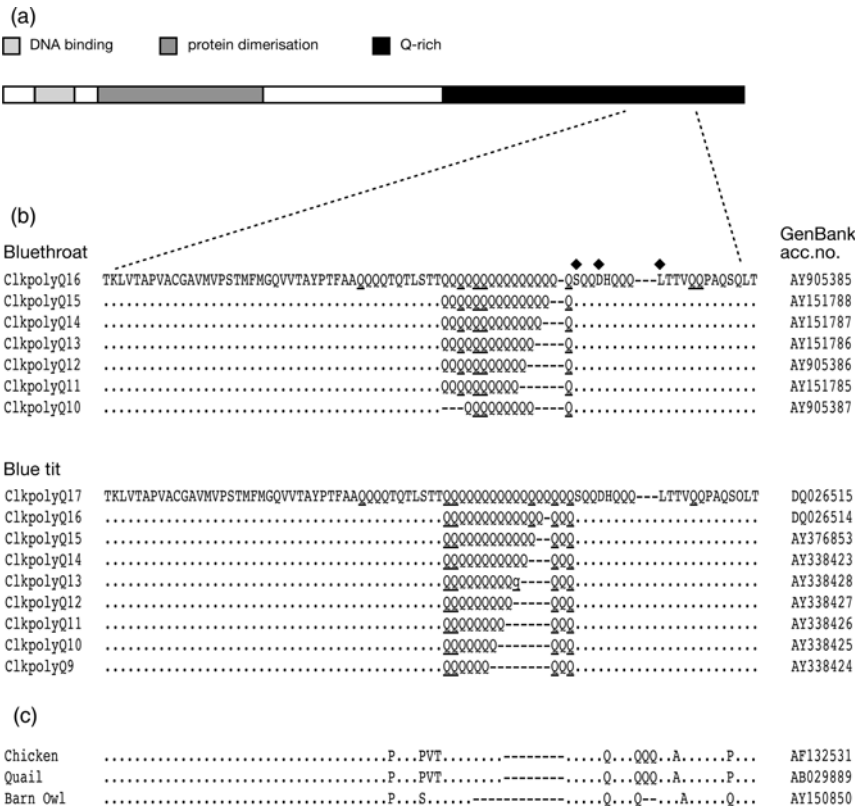


Fig. 2 Alignment of avian predicted CLOCK polyQ region sequences. (a) Schematic of the vertebrate CLOCK protein structure showing the DNA binding, protein dimerization and glutamine (Q) rich domains. (b) Predicted protein sequences of 16 passerine *ClkpolyQ*cds alleles. Alleles were named on the basis of the number of glutamine (Q) residues encoded by the variable length region. The predicted protein sequences corresponding to all 16 alleles are identical apart from polyglutamine repeat length variation. (c) Predicted CLOCK polyQ region sequences from published chicken, quail and barn owl *Clock* gene cDNA sequences. Gaps introduced to achieve alignment of all 19 avian CLOCK polyQ region sequences are indicated by (-); positions of sequence identity with the bluethroat *ClkpolyQ*₁₆ and blue tit *ClkpolyQ*₁₇ alleles are indicated by (.); evolutionarily conserved serine (S), aspartic acid (D) and leucine (L) residues used to assist sequence alignment are indicated by (♦). Passerine glutamine residues encoded by CAA are single-underlined while the lower-case Q within the blue tit *ClkpolyQ*₁₃ sequence is encoded by either CAA or CAG. The GenBank Accession nos associated with the sequences are shown.

iterations. Geographical distances between populations were calculated from their latitude and longitude, using an online geographical distance calculator at <http://snip-tools.com/latitudeLongitude.php>. We used *SPSS* 11.0 to perform standard parametric and nonparametric tests and to test whether the variables used were normally distributed. In the analyses of clinal variation, we present Pearson correlations (r) if the assumption of normality was met and Spearman rank correlations (r_s) otherwise. Following Weeks *et al.* (2006), we tested whether contrasts in mean allele size and the frequency of the two most common alleles were associated with geographical, latitudinal and/or longitudinal distance, using Mantel tests.

Results

Identification of passeriform *Clock* polyglutamine coding region polymorphism

Initially, *ClkpolyQ*cds were amplified from two blue tit genomic DNA samples using PCR primers that were designed based on evolutionarily conserved sequences shared among the reported vertebrate *Clock* gene sequences. Subsequently, primers were designed from these blue tit sequences and used to screen both bluethroat and blue tit genomic DNA samples for *ClkpolyQ*cds alleles that differed in length by increments of 3 bp. Seven bluethroat

and nine blue tit *ClkpolyQ*cds alleles were identified and their sequences determined from representative genomic DNA samples. Where possible, each *ClkpolyQ*cds allele sequence was obtained from a minimum of two individuals from two different populations. An alignment of the predicted protein sequences corresponding to the 16 passerine *ClkpolyQ*cds alleles is shown in Fig. 2. The amplified sequences were confirmed as being derived from avian *Clock* orthologues by their close homology with established avian *Clock* gene sequences (chicken, AF132531; quail, AB029889; barn owl, AY150850) (Fig. 2c). The predicted protein sequences corresponding to the 16 passerine *ClkpolyQ*cds sequences are identical except with respect to poly Q repeat length. At the nucleotide level, the alleles appear to be related by simple variation in poly CAG (i.e. poly Q encoding) repeat length (Fig. 2b). However, there are exceptions to this generalization; the simplest explanation for the origin of the rare bluethroat *ClkpolyQ*₁₀ allele (single allele found in the Heimdalen, Norway population, Table 1) is that it was derived from the *ClkpolyQ*₁₃ allele by deletion of the three most N-terminal of the poly Q codons in the variable repeat (Fig. 2b). Besides the poly Q coding region length polymorphism, few single nucleotide polymorphisms (SNP), defined as variants observed in a minimum of two independent amplification products, were found. Nonetheless, the bluethroat *ClkpolyQ*₁₁ allele had a synonymous G/A substitution located at position

37 of AY151785, while the blue tit *ClkpolyQ*₁₃ allele also exhibited an SNP with one of the variable length poly Q repeat glutamines being encoded by either CAA or CAG (Fig. 2b).

Population frequencies of the ClkpolyQcfs and microsatellite alleles

ClkpolyQcfs allele frequencies were determined for 12 bluethroat and 14 blue tit populations distributed throughout Europe and western Asia (Fig. 1, Table 1). The bluethroat populations displayed significantly lower levels of *ClkpolyQcfs* genetic variability (mean observed heterozygosity = 0.213, SEM = 0.039) than those of the blue tits (mean observed heterozygosity = 0.489, SEM = 0.040; Mann–Whitney *U*-test, *U* = 14.5, *n*₁ = 12, *n*₂ = 14, *P* < 0.001). This result reflects a single bluethroat *ClkpolyQcfs* allele (*ClkpolyQ*₁₃) being much more abundant than the other six bluethroat *ClkpolyQcfs* alleles (Table 1A). In contrast, four of the nine blue tit *ClkpolyQcfs* alleles (*ClkpolyQ*_{11–14}) were found at medium to high abundance (Table 1B).

Population *ClkpolyQcfs* allele frequencies were tested for deviation from Hardy–Weinberg equilibrium (HWE) omitting from the analysis two *ClkpolyQcfs* monomorphic (*H* = 0.000) populations (bluethroat, Valduerna; blue tit, Pantelleria; Table 1), and one population with a very low level (*H* = 0.048) of allelic variation (bluethroat, Stugudal; Table 1A). Among the 10 bluethroat populations examined, *ClkpolyQcfs* allele frequencies did not deviate significantly from HWE, either overall (*P* = 0.84) or within populations (for all populations *P* > 0.12). In contrast, the blue tit populations deviated significantly from HWE at the *ClkpolyQcfs* locus, both overall (*P* = 0.0002) and within 4 of the 13 populations tested (Jomfruland, Vosbergen, Antwerp and Tübingen; all *P* < 0.047), with one additional population showing a similar but nonsignificant tendency (Muro, *P* = 0.057). All of the five blue tit populations showing evidence of deviation from HWE displayed a deficiency of heterozygotes (mean *F*_{IS} = 0.09, range 0.05–0.16).

Equivalent analyses based on allele frequencies at nine polymorphic microsatellite loci revealed that of the five blue tit populations showing signs of *ClkpolyQcfs* heterozygote deficiency, only the Jomfruland population showed a consistent pattern across the microsatellite loci. Five of nine microsatellite loci displayed a significant deficiency of heterozygotes in the Jomfruland population and eight of nine *F*_{IS} values were positive (mean *F*_{IS} = 0.13, range –0.01–0.43). This may be due to some degree of inbreeding in this island population, located about 3.5 km off the coast of south-eastern Norway. In the other populations, heterozygote deficiency observed at only one or a few markers could result from null alleles or allelic dropout, although none of the microsatellite markers used have displayed such problems in previous analyses of blue tit family data

(Delhey *et al.* 2003; B. Kempenaers, unpublished data). The mean *F*_{IS} values, calculated from nine microsatellite loci for the other four *ClkpolyQcfs* heterozygosity deficient populations (Vosbergen, Antwerp, Tübingen and Muro) were low (range: 0.014–0.025), and similar to those of the eight populations that did not show signs of deviation from HWE at the *ClkpolyQcfs* locus (range –0.024–0.031). The Pantelleria population showed no significant departure from HWE across the microsatellite loci (*P* = 0.17), but it did exhibit much lower levels of heterozygosity than the other blue tit populations (Pantelleria: mean *H* = 0.40; remaining 13 populations: mean *H* = 0.81, range: 0.73–0.86). This is due to a much lower allelic diversity at all eight microsatellite loci that amplified in samples from this population (Pantelleria, mean number of alleles = 3.1; remaining 13 populations, mean number of alleles = 18.1; range, 15.1–24.6).

Interpopulation comparisons of ClkpolyQcfs and microsatellite allele frequencies

Intraspecific *F*_{ST} values, combining all populations, revealed significant population differentiation, at the *ClkpolyQcfs* locus, for both species (bluethroat: *F*_{ST} = 0.053, *P* = 0.0001; blue tit: *F*_{ST} = 0.045, *P* = 0.0001; all *P* values based on 10 000 permutations). This result was supported by pairwise comparisons (Table S1). The 66 pairwise comparisons among bluethroat populations yielded six *F*_{ST} values that remained statistically significant after applying a table-wide Bonferroni correction for multiple tests, while 22 comparisons showed a *P* < 0.05 (Table S1A). Among the 91 blue tit pairwise comparisons, 28 *F*_{ST} values were significantly different from zero after Bonferroni correction, while 53 comparisons had *P* < 0.05 (Table S1B). After excluding the Pantelleria population, monomorphic at the *ClkpolyQcfs* locus, from the blue tit analysis, the level of *ClkpolyQcfs* locus differentiation decreased but remained statistically significant (*F*_{ST} = 0.018, *P* = 0.0001). For both species, the overall level of genetic differentiation was similar when based on microsatellite allele frequencies (bluethroat: *F*_{ST} = 0.040, *P* = 0.0002; Johnsen *et al.* 2006; blue tit: all 14 populations, *F*_{ST} = 0.040, *P* = 0.0001; excluding Pantelleria, *F*_{ST} = 0.017, *P* = 0.0001).

In the bluethroat, no significant association was found between interpopulation geographical distance and levels of *ClkpolyQcfs* allelic differentiation as measured by *F*_{ST} (Mantel test: *r* = –0.12, *P* = 0.28). In contrast, the blue tit *F*_{ST} values indicated a weak tendency towards isolation by distance when including the *ClkpolyQcfs* monomorphic Pantelleria population (Mantel test: *r* = 0.38, *P* = 0.084) but this tendency disappeared when the Pantelleria data were excluded (Mantel test: *r* = 0.13, *P* = 0.15). Similar results were found when looking for an association between interpopulation geographical distances and population

differentiation measured using microsatellite F_{ST} values [bluethroat, Mantel test: $r = 0.10$, $P = 0.29$; blue tit 14 populations (i.e. with Pantelleria), Mantel test: $r = 0.35$, $P = 0.091$; blue tit 13 populations (i.e. without Pantelleria), Mantel test: $r = 0.17$, $P = 0.19$].

In both bird species, initial analyses suggested some phylogenetic structuring with respect to *ClkpolyQc*s allele frequency differentiation (as measured by F_{ST} values): within-subspecies comparisons revealed relatively low levels of allelic differentiation (bluethroat: $n = 11$, mean $\pm$ SE $F_{ST} = 0.010 \pm 0.010$; blue tit: $n = 46$, mean $\pm$ SE $F_{ST} = 0.020 \pm 0.003$), while greater levels of differentiation were found in between-subspecies comparisons (bluethroat: $n = 55$, mean $\pm$ SE $F_{ST} = 0.059 \pm 0.011$; blue tit: $n = 45$, mean $\pm$ SE $F_{ST} = 0.100 \pm 0.021$) (Table S1). These F_{ST} differences were statistically significant in the blue tit ($U = 755.5$, $P = 0.027$), but not in the bluethroat ($U = 207.0$, $P = 0.10$). However, the difference in blue tits was highly influenced by the inclusion of the Pantelleria population, which is clearly distinct from all other 13 blue tit populations at the *ClkpolyQc*s locus (Table S1). This result reflects the Pantelleria population being monomorphic, having only the *ClkpolyQ*₁₂ allele, a finding consistent with Pantelleria being a genetically isolated and distinct population. This is corroborated by F_{ST} analyses based on microsatellites, where all pairwise comparisons involving this population were highly significant (data not shown). When excluding the Pantelleria population from the blue tit *ClkpolyQc*s analyses, there was no significant difference in F_{ST} -values calculated from within-subspecies ($n = 46$, mean $F_{ST} = 0.020 \pm 0.004$) and between-subspecies comparisons ($n = 32$, mean $F_{ST} = 0.016 \pm 0.004$; $U = 716.5$, $P = 0.84$).

In contrast, the analysis of the blue tit microsatellite data revealed a significant phylogenetic signal both when including Pantelleria (within-subspecies: $n = 46$, mean $F_{ST} = 0.007 \pm 0.001$; between subspecies: $n = 45$, mean $F_{ST} = 0.111 \pm 0.019$, $U = 243.5$, $P < 0.001$) and when excluding it (within-subspecies: $n = 46$, mean $F_{ST} = 0.008 \pm 0.001$; between subspecies: $n = 32$, mean $F_{ST} = 0.033 \pm 0.003$, $U = 221.5$, $P < 0.001$). Supporting evidence that the microsatellite allele frequencies consistently reflected population/phylogenetic structure comes from Mantel tests of the relationships between F_{ST} values based on single microsatellite markers and those based on the remaining markers, which revealed significant correlations for eight out of nine markers (Mantel tests: mean $r = 0.80$, range = 0.50–0.93, all $P < 0.016$; Pantelleria excluded). Only PC8 did not show a significant pattern (Mantel test: $r = 0.32$, $P = 0.13$), which may partly be explained by the high number of alleles at this locus (76 alleles) making it less suitable as a phylogenetic marker. A similar pattern of population/phylogenetic structuring was also evident in the bluethroat microsatellite allele frequency data (see Johnsen et al. 2006).

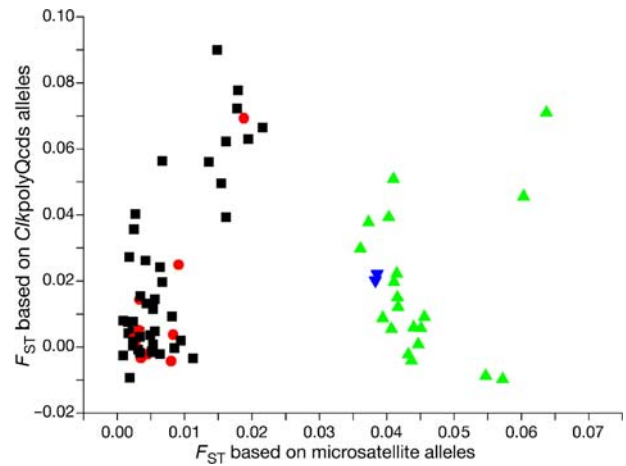


Fig. 3 Relationships between blue tit interpopulation F_{ST} values calculated from microsatellite and *ClkpolyQc*s allele frequencies. Allele frequencies from nine microsatellite loci and the *ClkpolyQc*s locus were used to calculate pairwise F_{ST} values for 13 blue tit populations, excluding Pantelleria. Within- and between-subspecies comparisons are distinguished: black squares, within-subspecies comparisons; red circles, *caeruleus* vs. *obscurus*; green up-triangles, *caeruleus* vs. *ogliastrea*; blue down-triangles, *obscurus* vs. *ogliastrea*.

Among the blue tit populations, there was a significant association between pairwise F_{ST} values based on *ClkpolyQc*s allele frequency and those based on microsatellite allele frequencies when including Pantelleria (Mantel test: $r = 0.95$, $P = 0.007$), but no association when excluding Pantelleria (Mantel test: $r = 0.19$, $P = 0.18$; Fig. 3). In the bluethroat, the two sets of pairwise F_{ST} values were significantly associated (Mantel test: $r = 0.61$, $P = 0.009$).

Latitudinal distribution of *ClkpolyQc*s and microsatellite allele frequencies

Given the role of the *Clock* gene in the vertebrate circadian oscillator, we hypothesized that a relationship might exist between *ClkpolyQc*s allele frequencies and latitude. Among the 12 bluethroat populations, no significant correlation was found between population average *ClkpolyQc*s allele length and latitude ($n = 12$, $r = -0.28$, $P = 0.39$). In contrast, among the 14 blue tit populations, there was a statistically significant correlation between average allele length and latitude ($n = 14$, $r = 0.64$, $P = 0.013$), with northern populations having a longer mean allele length than more southern populations (Fig. 4). When the Pantelleria population was excluded from the analysis, a positive, albeit statistically nonsignificant, correlation was found ($n = 13$, $r = 0.46$, $P = 0.11$). No correlations were found between average *ClkpolyQc*s allele size and longitude in either bird species (blue tit: $n = 14$, $r = 0.19$, $P = 0.53$; bluethroat: $n = 12$, $r_s = -0.07$, $P = 0.82$). We found similar relationships between average allele length and latitude after restricting the analyses to the 1303 blue tit samples

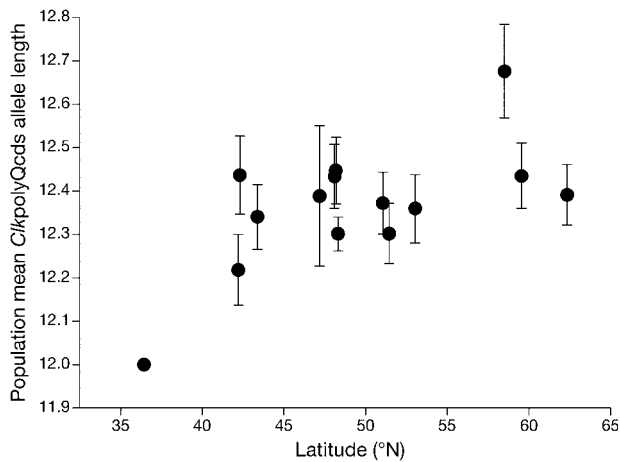


Fig. 4 Mean ($\pm$ SEM) *ClkpolyQcde* allele length within 14 blue tit populations plotted as a function of population breeding latitude.

from which we obtained genotypes for both *ClkpolyQcde* and four or more microsatellite markers (including Pantelleria: $n = 14$, $r = 0.63$, $P = 0.015$; excluding Pantelleria: $n = 13$, $r = 0.45$, $P = 0.12$).

To explore the blue tit *ClkpolyQcde* allelic frequency spatial patterns in more detail, we performed Mantel tests on contrasts in mean allele length, and the frequency of the two most common *ClkpolyQcde* alleles against geographical, latitudinal and longitudinal distance. We conservatively excluded the Pantelleria population from these, and all subsequent analyses, because our previous analyses revealed that this population is genetically distinct with respect to *ClkpolyQcde* and may well be genetically completely isolated from continental European blue tit populations. It should be noted, however, that all patterns described with respect to distance and latitude are strengthened by including Pantelleria data (data not shown). There was a significant positive association between contrasts in mean blue tit *ClkpolyQcde* allele length and latitudinal distance, but none for either geographical or longitudinal distance (Table 2). The frequency of the most common blue tit *ClkpolyQcde* allele (*ClkpolyQ*₁₂) was significantly negatively related to latitude, while the association was weakly positive for the second most common allele (*ClkpolyQ*₁₃), although not significantly so (Table 2).

As there was evidence of a cline in blue tit but not bluethroat mean *ClkpolyQcde* allele length, we restricted equivalent analyses of microsatellite data to the blue tits. Contrasts in marker-wide mean microsatellite allele length tended to be negatively associated with latitudinal distance, although not significantly so [Table 2; only individuals typed for all nine microsatellite loci ($n = 1127$) included]. Looking at the mean allele length of individual microsatellite loci, PC7 showed a significant positive association with latitude, PK12 a significant negative association and PC3 a negative trend (Table 2). For the remaining six microsatel-

lite loci, there were no relationships with latitude (all $|r| < 0.18$, all $P > 0.15$). For both the PC7 and PK12 loci, mean allele length was similarly related to both geographical and latitudinal distance (Table 2). Furthermore, the frequencies of both the most and second most common PC7 allele showed significant associations with both geographical and latitudinal distance, while the most common PK12 allele was significantly related to latitude only (Table 2). Only one statistically significant relationship was found between microsatellite characteristics and longitude, with eastern blue tit populations having longer marker-wide mean microsatellite allele length than more western ones (Table 2).

There was no consistent pattern of linkage disequilibrium between any of the three microsatellite loci that showed evidence of clinal variation (PC7, PK12, PC3), and the *ClkpolyQcde* locus across all populations (data not shown).

Average *ClkpolyQcde* allele size is perhaps a crude summary statistic for population characterization and its use could well obscure interesting genetic features. We hypothesized that *ClkpolyQ* alleles were likely to be codominant at the molecular level because of the CLOCK protein's mode of action, that of dimerizing with the BMAL1 protein before binding to target control elements in the genome. Therefore, we classified blue tit genotypes into one of three categories: (1) birds having two alleles *ClkpolyQ*₁₂ or smaller ($\leq Q_{12}/\leq Q_{12}$), (2) heterozygous birds with one allele *Q*₁₂ or smaller and the other *Q*₁₃ or larger ($\leq Q_{12}/\geq Q_{13}$), and (3) birds with both alleles *Q*₁₃ or larger ($\geq Q_{13}/\geq Q_{13}$) (Table 3). The selection of the *Q*₁₂/*Q*₁₃ boundary was based on these allele lengths lying in the middle of the observed blue tit *ClkpolyQ* allele size range (*Q*₉–*Q*₁₇) (Table 1B). Furthermore, *Q*₁₂ allele frequency was negatively associated with latitude while *Q*₁₃ allele frequency was positively associated, albeit not statistically significantly so (Table 2). The proportion of category 3 ($\geq Q_{13}/\geq Q_{13}$) birds was positively correlated with latitude ($n = 13$, $r = 0.67$, $P = 0.012$), and this remained so when limiting the analysis to the 10 blue tit populations belonging to the subspecies *caeruleus* (see Table 1B) ($n = 10$, $r_s = 0.75$, $P = 0.013$). No significant correlations were found between the proportions of category 1 ($\leq Q_{12}/\leq Q_{12}$) and category 2 ($\leq Q_{12}/\geq Q_{13}$) birds, and latitude (both $P > 0.21$) or between the proportions of any of the three genotype categories and longitude (all $P > 0.22$). When comparing the relationships between contrasts in proportion of category 3 birds, and latitudinal, longitudinal and geographical distance, respectively, we found a stronger association with latitude (Mantel test, $r = 0.49$, $P = 0.001$) than with distance (Mantel test, $r = 0.37$, $P = 0.026$), and no association with longitude (Mantel test, $r = 0.17$, $P = 0.15$).

Analogous analyses of the bluethroat data set (categorizing genotypes as 1: $\leq Q_{13}/\leq Q_{13}$, 2: $\leq Q_{13}/\geq Q_{14}$ or 3: $\geq Q_{14}/\geq Q_{14}$) revealed no significant correlations with either latitude or longitude (all $P > 0.37$).

Table 3 Number (percentages) of individuals in 14 blue tit populations belonging to each of three genotype categories: 1, having two alleles *Clk*polyQ₁₂ or smaller ($\leq Q_{12}/\leq Q_{12}$); 2, heterozygous birds with one allele Q₁₂ or smaller and the other Q₁₃ or larger ($\leq Q_{12}/\geq Q_{13}$); and 3, birds with both alleles Q₁₃ or larger ($\geq Q_{13}/\geq Q_{13}$)

Population	<i>n</i>	Category 1 ($\leq Q_{12}/\leq Q_{12}$)	Category 2 ($\leq Q_{12}/\geq Q_{13}$)	Category 3 ($\geq Q_{13}/\geq Q_{13}$)
Jyväskylä, Finland	97	35 (36.1)	46 (47.4)	16 (16.5)
Oslo, Norway	77	35 (45.5)	30 (39.0)	12 (15.6)
Jomfruland, Norway	85	27 (31.8)	35 (41.2)	23 (27.1)
Vosbergen, Netherlands	96	52 (54.2)	29 (30.2)	15 (15.6)
Wytham Woods, UK	96	42 (43.8)	42 (43.8)	12 (12.5)
Antwerp, Belgium	102	43 (46.1)	46 (42.2)	13 (11.8)
Tübingen, Germany	285	142 (49.8)	115 (40.4)	28 (9.8)
Forêt d'Orient, France	103	39 (37.9)	49 (47.6)	15 (14.6)
Vienna, Austria	91	27 (29.7)	50 (54.9)	14 (15.4)
Auxonne, France	18	10 (55.6)	6 (33.3)	2 (11.1)
Rouviere, France	91	36 (39.6)	46 (50.5)	9 (9.9)
Muro, Corsica, France	95	45 (47.4)	37 (38.9)	13 (13.7)
Pirio, Corsica, France	96	54 (56.3)	36 (37.5)	6 (6.3)
Pantelleria, Italy	90	90 (100)	0 (0)	0 (0)

n, number of *Clock*-genotyped individuals.

Discussion

Clock gene sequence variation

Repetitive DNA sequences are believed to have relatively high mutation rates because of 'slippage' during replication (Wren *et al.* 2000). Consequently, the existence of length variation in the passerine *Clock* poly Q coding region could be attributed to high mutation rates combined with relaxed selection. Alternatively, the *Clock* poly Q region may be under balancing selection, with genetic diversity being sustained by spatial or temporal heterogeneity in selection pressures. At the molecular genetic level, three observations are pertinent to the issue of whether or not the observed *Clock* gene allelic variation is selectively neutral or rather maintained by some form of balancing selection. First, the *CLOCK* protein poly Q repeat, along with its flanking sequences, is conserved within class Aves and this conservation appears to limit poly Q repeat length variation (Fidler & Gwinner 2003; this study). Such evolutionary sequence conservation is generally considered indicative of functional constraints (Nei & Kumar 2000). The second observation is simply that poly Q repeat length polymorphism is maintained in different avian species, albeit in differing ratios (Fidler & Gwinner 2003; this study). The scarcity of analogous genetic variability in the human *Clock* gene suggests that such genetic variation is not an inevitable and selectively neutral consequence of the vertebrate *Clock* gene structure (Saleem *et al.* 2001). The third observation is that genetic code redundancy makes it entirely feasible to encode poly Q repeats of stable length, simply by mixing CAA and CAG codons and thereby inhibiting 'slippage'

during DNA replication (Wren *et al.* 2000; Choudhry *et al.* 2001; Fondon & Garner 2004; reviewed in Hancock & Simon 2005). Indeed, such alternative glutamine codons are present among the N and C termini codons of the passeriform *Clock* poly Q coding region (Fig. 2). That selection has not stabilized the *Clock* poly CAG repeat at a single length is at least suggestive that the variation is itself selectively advantageous (Wren *et al.* 2000).

Contrasting population genetics of the ClkpolyQcfs and microsatellite loci

Population structuring arising from phylogenetic history and nonadaptive evolutionary processes (e.g. founder effects and genetic drift) complicates comparative population genetic studies of putatively functional genetic variation (Nei & Kumar 2000). To address this issue, we contrasted interpopulation comparisons at the *Clk*polyQcfs locus with those at microsatellite loci.

The Pantelleria blue tit population appears genetically distinct, as it is monomorphic for the *Clk*polyQ₁₂ allele, has relatively low levels of microsatellite allelic variation and is strongly differentiated from all other blue tit populations as measured using both *Clk*polyQcfs and microsatellite allele frequencies. The phylogenetic relationships between the Pantelleria and continental European blue tits are somewhat unclear. Pantelleria birds have been placed in the *ultramarinus* subspecies based on morphological appearance (Cramp & Perrins 1993). Furthermore, two molecular phylogenetic studies, using mitochondrial DNA sequences, have placed North African specimens of *ultramarinus* closer to the Canary Island subspecies than to the European

blue tits (Salzburger *et al.* 2002; Kvist *et al.* 2005). The monomorphism of the *ClkpolyQcds* locus in the Pantelleria sample could be a result of strong selection for a particular allele and/or a consequence of the geographical isolation, and phylogenetic history, of this small island population. Therefore, erring on the side of caution, in the remainder of this discussion, we only refer to results from analyses that excluded the Pantelleria population data.

There was no difference in the overall level of genetic differentiation of blue tit populations on the basis of either *ClkpolyQcds* or microsatellite allele frequencies. However, two lines of evidence suggest that interpopulation variation in *ClkpolyQcds* allele frequencies is not simply a reflection of demographic processes among blue tit populations. First, there was a strong phylogenetic signal in the microsatellite data, which was not the case, in the absence of the Pantelleria data, for *ClkpolyQcds* allele frequency variation. Second, pairwise F_{ST} values based on *ClkpolyQcds* were not significantly associated with those based on microsatellites, showing that they do not reflect similar patterns of population structure and/or history (Fig. 3). In contrast, for eight of the nine microsatellite markers, single-marker F_{ST} values were highly correlated with those calculated from the remaining markers, showing that allele frequencies of these markers were collectively related through underlying patterns of demography and population structure.

There was some evidence of deviation from HWE expectations in the blue tit *ClkpolyQcds* allele frequencies, with a deficiency of heterozygotes in five out of the 13 populations tested. Such deviations of genotype frequencies from HWE could reflect any of the HWE assumptions not being met (e.g. 'founder' effects, nonrandom mating, inbreeding, selection). However, it is noteworthy that heterozygote deficiency is consistent with Wahlund's principle; that in populations subdivided into smaller breeding units, the frequency of homozygotes tends to be higher than predicted from HWE (Nei & Kumar 2000). In contrast to the *ClkpolyQcds* locus, there was little evidence of deviation from HWE among the nine blue tit microsatellite loci.

Thus, the accumulated evidence indicates that among the blue tit but not bluethroat populations examined, the *ClkpolyQcds* locus shows patterns of allele distribution that do not simply reflect those of putatively selectively neutral microsatellite loci.

ClkpolyQcds allele and genotype frequencies in relation to latitude

We hypothesized that, given the *Clock* gene's role in both the circadian clock core mechanism and its output, and the role of the circadian clock in interpreting photoperiod (Schultz & Kay 2003; Hazlerigg & Wagner 2006), latitudinal clines in *ClkpolyQcds* allele frequencies may exist, reflecting local adaptation to latitudinal gradients in photoperiod

parameters such as seasonal rate-of-change of photoperiod. Indeed, among the blue tit, but not bluethroat populations, we found evidence of latitudinal clines in both *ClkpolyQcds* allele and genotype frequencies, with longer *ClkpolyQcds* alleles being more common at higher latitudes. Among the microsatellites, two (PC7, PK12) of nine markers showed significant associations with latitude in several parameters, albeit in opposite directions, while the mean allele size of a third marker (PC3) was almost significantly related to latitude. It should be noted that latitudinal distance was strongly correlated with geographical distance (Mantel test: $r = 0.92$, $P = 0.0003$), making it difficult to distinguish the relative significance of these two spatial parameters. Both of the microsatellite markers that showed significant clinal variation with latitude (PC7, PK12) displayed a similar association with geographical distance (Table 2). This was not the case with the cline in mean *ClkpolyQcds* allele size, which was significant for latitude but not for geographical distance (Table 2). Similarly, the frequency of the *ClkpolyQcds* category 3 genotype ($\geq Q_{13}/\geq Q_{13}$) was more strongly related to latitude than to distance. Given the low sample sizes ($n = 13$ populations) and the resulting low power of some of these tests, we obviously cannot make strong inferences regarding the direction of the cline in *ClkpolyQcds* alleles. Nevertheless, we note that *ClkpolyQcds* allele frequencies correlate more strongly with latitude while variation in PC7 and PK12 microsatellite allele frequencies are equally well described by distance per se, and could arise from geographically restricted gene flow (Vasemägi 2006). Importantly, the latitudinal cline in the frequency of category 3 ($\geq Q_{13}/\geq Q_{13}$) blue tits persisted when confining the analysis to the 10 *caeruleus* subspecies populations. Hence, the positive correlation between the proportion of individuals with long alleles and latitude appears robust and not simply a result of phylogenetic constraints. We found no evidence of linkage disequilibrium between the three microsatellites showing signs of clinal variation (PC3, PC7, PK12) and the *ClkpolyQcds* locus, but we cannot exclude the possibility that clinal variation in the microsatellite markers is due to linkage disequilibrium with other functional genes under latitude-related selection (see Gockel *et al.* 2001 for a similar example in *Drosophila melanogaster*).

The correlative nature of this study obviously limits the strength of any inferences that can be made regarding selection on *ClkpolyQcds* allele length as the cause of the observed correlations. However, a number of behaviours are expected to correlate with latitude, including the timing of reproduction and the occurrence, and timing, of dispersal/migration. We note that blue tits are partially migratory in the northern parts of their distribution, with a fraction of the population dispersing/migrating during autumn while the majority are year-round residents (Cramp & Perrins 1993). A possibility that should be

explored further is whether longer *ClkpolyQc*ds alleles are associated with the propensity to migrate or disperse. The absence of a latitudinal cline in bluethroat *ClkpolyQc*ds allele frequencies may reflect all bluethroats being migratory and therefore not so strongly influenced by the photoperiodic and/or climatic parameters of specific latitudes when compared with the more resident blue tits. Similarities between patterns of variation in microsatellite markers and *ClkpolyQc*ds alleles among the bluethroat populations suggest that, in this species, allelic variation at the *ClkpolyQc*ds locus is not subjected to detectable levels of selection. However, the lack of statistical significance in the bluethroat data may also reflect the smaller sample sizes for this species when compared with the blue tit both in terms of number of populations (12 vs. 14) and average number of individuals sampled per population (31 vs. 102).

Phenotypic consequences of tandem repeat length polymorphisms

Tandem repeat length polymorphisms, within both coding and control regions, are associated with both subtle and extreme phenotypic effects (La Spada & Taylor 2003; Zitzmann & Nieschlag 2003; Fondon & Garner 2004; Hammock & Young 2005). Indeed it has been hypothesized that the relatively high mutation rates of repeat sequences may account for rapid morphological evolution among mammals (Fondon & Garner 2004). The potential of coding region repeats for rapid evolution might even be selected for, as it provides plasticity in the face of fluctuating selective pressures (Wren *et al.* 2000). Clearly similar arguments can be forwarded for the possible evolutionary significance of tandem repeats within the coding, and control, regions of behaviour-related genes (Hammock & Young 2005). Determination of the phenotypic effects of the different *ClkpolyQc*ds alleles described here would require detailed studies of both circadian and photoperiod-related behaviours of birds of differing *ClkpolyQ* genotypes.

Summary

In summary, we have described, in two Eurasian passerine species, multiple alleles of the avian *Clock* gene varying in the length of a repeat sequence encoding a polyglutamine tract. Among blue tit, but not bluethroat populations, there is evidence suggesting that *ClkpolyQc*ds allele frequencies reflect different processes than those acting on supposedly neutral microsatellites. However, our data do not permit any firm conclusions regarding whether *Clock* gene allelic variation is under current selection or not. Further study of the geographical and temporary patterns of *ClkpolyQc*ds alleles, along with descriptions of behavioural differences associated with different *ClkpolyQc*ds genotypes, may

help to delineate possible selective pressures maintaining this genetic polymorphism. More generally, we propose that investigating changing frequencies of allelic variants of genes encoding circadian clock components may warrant attention in the context of adaptation to rapid climate change. Finally, we support the growing view that population genetic studies of behaviour-related genes in free-living, nonmodel, animals may provide valuable insights into both the evolution and functions of such genes.

Acknowledgements

The following people are thanked for generously collecting and sharing bluethroat blood samples: Martin S. Adamian, Bob Chutny, Javier García Fernández, Dieter Franz, Geir Grimsby, Per Øyvind Grimsby, Jan T. Lifjeld, Václav Pavel, Anatoly Ostaschenko, Sophie Questiau and Michael Räss. Jan T. Lifjeld is also thanked for providing blue tit DNA samples from the avian DNA/tissue collection of the Natural History Museum, University of Oslo, and Gabriele Sorci kindly provided samples from the Auxonne and Pantelleria blue tit populations. Thanks also go to the 'Montpellier team' for their continuous efforts in the study of Mediterranean blue tit populations, to Mihai Valcu for help with statistical analyses, to Staffan Bensch and three anonymous reviewers for constructive comments and to Theo Weber for graphical assistance. This study was supported by the Max Planck Society, the Alexander von Humboldt foundation, the Norwegian Research Council, Parsyst (European Union) and the National Centre for Biosystematics (University of Oslo).

References

- Avivi A, Albrecht U, Oster H *et al.* (2001) Biological clock in total darkness: the Clock/MOP3 circadian system of the blind subterranean mole rat. *Proceedings of the National Academy of Sciences, USA*, **98**, 13751–13756.
- Bensch S, Price T, Kohn J (1996) Isolation and characterization of microsatellite loci in a *Phylloscopus* warbler. *Molecular Ecology*, **5**, 150–151.
- Boake CRB, Arnold SJ, Breden F *et al.* (2002) Genetic tools for studying adaptation and the evolution of behavior. *American Naturalist*, **160**, S143–S159.
- Choudhry S, Mukerji M, Srivastava AK, Jain S, Brahmachari SK (2001) CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. *Human Molecular Genetics*, **10**, 2437–2446.
- Cramp S, Perrins CM (1993) Blue tit. In: *The Birds of the Western Palearctic* (eds Cramp S, Perrins CM), pp. 225–248. Oxford University Press, Oxford, UK.
- Dawson DA, Hanotte O, Greig C, Stewart IRK, Burke T (2000) Polymorphic microsatellites in the blue tit *Parus caeruleus* and their cross-species utility in 20 songbird families. *Molecular Ecology*, **9**, 1941–1944.
- Delhey K, Johnsen A, Peters A, Andersson S, Kempnaers B (2003) Paternity analysis reveals opposing selection pressures on crown colouration in the blue tit (*Parus caeruleus*). *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 2057–2063.
- Double MC, Dawson D, Burke T, Cockburn A (1997) Finding the fathers in the least faithful bird: a microsatellite-based genotyping

- system for the superb fairy-wren *Malurus cyaneus*. *Molecular Ecology*, **6**, 691–693.
- Fidler AE, Gwinner E (2003) Comparative analysis of avian BMAL1 and CLOCK protein sequences: a search for features associated with owl nocturnal behaviour. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, **136**, 861–874.
- Fitzpatrick MJ, Ben-Shahar Y, Smid HM *et al.* (2005) Candidate genes for behavioural ecology. *Trends in Ecology & Evolution*, **20**, 96–104.
- Fondon JW, Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. *Proceedings of the National Academy of Sciences, USA*, **101**, 18058–18063.
- Gockel J, Kennington WJ, Hoffmann A, Goldstein DB, Partridge L (2001) Nonclinality of molecular variation implicates selection in maintaining a morphological cline of *Drosophila melanogaster*. *Genetics*, **158**, 319–323.
- Goodman SJ (1997) RSTCALC: a collection of computer programs for calculating unbiased estimates of genetic differentiation and determining their significance for microsatellite data. *Molecular Ecology*, **6**, 881–885.
- Goudet J (1995) FSTAT, version 1.2: a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.
- Hammock EAD, Young LJ (2005) Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science*, **308**, 1630–1634.
- Hancock JM, Simon M (2005) Simple sequence repeats in proteins and their significance for network evolution. *Gene*, **345**, 113–118.
- Hazlerigg DG, Wagner GC (2006) Seasonal photoperiodism in vertebrates: from coincidence to amplitude. *Trends in Endocrinology and Metabolism*, **17**, 83–91.
- Iuvone PM, Tosini G, Pozdeyev N *et al.* (2005) Circadian clocks, clock networks, arylalkylamine N-acetyltransferase, and melatonin in the retina. *Progress in Retinal and Eye Research*, **24**, 433–456.
- Johnsen A, Andersson S, Fernandez JG *et al.* (2006) Molecular and phenotypic divergence in the bluethroat (*Luscinia svecica*) subspecies complex. *Molecular Ecology*, **15**, 4033–4047.
- Katzberg D, Young T, Finn L *et al.* (1998) A CLOCK polymorphism associated with human diurnal preference. *Sleep*, **21**, 569–576.
- Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite populations. *Genetics*, **148**, 1921–1930.
- Ko CH, Takahashi JS (2006) Molecular components of the mammalian circadian clock. *Human Molecular Genetics*, **15**, R271–R277.
- Kvist L, Broggi J, Illera JC, Koivula K (2005) Colonisation and diversification of the blue tits (*Parus caeruleus teneriffae*-group) in the Canary Islands. *Molecular Phylogenetics and Evolution*, **34**, 501–511.
- La Spada AR, Taylor JP (2003) Polyglutamines placed into context. *Neuron*, **38**, 681–684.
- Leder EH, Danzmann RG, Ferguson MM (2006) The candidate gene, *Clock*, localizes to a strong spawning time quantitative trait locus region in rainbow trout. *Journal of Heredity*, **97**, 74–80.
- Mishima K, Tozawa T, Satoh K, Saitoh H, Mishima Y (2005) The 311T/C polymorphism of hClock is associated with evening preference and delayed sleep timing in a Japanese population sample. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, **133**, 101–104.
- Mitchell PJ, Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, **245**, 371–378.
- Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Ohta T, Kimura M (1973) The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. *Genetical Research*, **22**, 201–204.
- Otter K, Ratcliffe L, Michaud D, Boag PT (1998) Do female black-capped chickadees prefer high-ranking males as extra-pair partners? *Behavioral Ecology and Sociobiology*, **43**, 25–36.
- Panda S, Hogenesch JB, Kay SA (2002) Circadian rhythms from flies to human. *Nature*, **417**, 329–335.
- Reppert SM, Weaver DR (2002) Coordination of circadian timing in mammals. *Nature*, **418**, 935–941.
- Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.
- Saleem Q, Anand A, Jain S, Brahmachari SK (2001) The polyglutamine motif is highly conserved at the *Clock* locus in various organisms and is not polymorphic in humans. *Human Genetics*, **109**, 136–142.
- Salzburger W, Martens J, Sturmbauer C (2002) Paraphyly of the blue tit (*Parus caeruleus*) suggested from cytochrome *b* sequences. *Molecular Phylogenetics and Evolution*, **24**, 19–25.
- Schultz TF, Kay SA (2003) Circadian clocks in daily and seasonal control of development. *Science*, **301**, 326–328.
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.
- Steeves TDL, King DP, Zhao YL *et al.* (1999) Molecular cloning and characterization of the human CLOCK gene: expression in the suprachiasmatic nuclei. *Genomics*, **57**, 189–200.
- Tauber E, Kyriacou CP (2005) Molecular evolution and population genetics of circadian clock genes. *Methods in Enzymology*, **393**, 797–817.
- Vasemägi A (2006) The adaptive hypothesis of clinal variation revisited: single-locus clines as a result of spatially restricted gene flow. *Genetics*, **173**, 2411–2414.
- Weeks AR, Mckechne SW, Hoffmann AA (2006) In search of clinal variation in the period and clock timing genes in Australian *Drosophila melanogaster* populations. *Journal of Evolutionary Biology*, **19**, 551–557.
- Weir B, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Wren JD, Forgacs E, Fondon JW *et al.* (2000) Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. *American Journal of Human Genetics*, **67**, 345–356.
- Young MW, Kay SA (2001) Time zones: a comparative genetics of circadian clocks. *Nature Reviews Genetics*, **2**, 702–715.
- Zitzmann M, Nieschlag E (2003) The CAG repeat polymorphism within the androgen receptor gene and maleness. *International Journal of Andrology*, **26**, 76–83.

Arild Johnsen is an evolutionary ecologist whose focus is on sexual selection, genetic mate choice and phylogeography. Andrew Fidler is a molecular biologist interested in the behavioural genetics of free living animals. Sylvia Kuhn, Kim Carter and Alexandra Hoffmann contributed via their laboratory skills. Bart Kempnaers is head of the Department of Behavioural Ecology and Evolutionary Genetics, and together with Andrew Fidler, initiated this work. His main interests are sexual selection and the evolution of mating systems. The other coauthors are ecologists who contributed samples from their study population of blue tits, and who commented on the manuscript.

Supplementary material

The following supplementary material is available for this article:

Table S1 Pairwise F_{ST} (above diagonal) and R_{ST} (below diagonal) values for (A) 12 bluethroat (*Luscinia svecica*) populations and (B) 14 blue tit (*Cyanistes caeruleus*) populations. Comparisons with P values < 0.05 are shown in bold. * indicates a value significant after table-wide Bonferroni correction [for (A): critical P value: 0.000758; for (B): critical P value: 0.000549; Rice 1989].

This material is available as part of the online article from:

<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03552.x>

(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.