

Search for informative polymorphisms in candidate genes: clock genes and circadian behaviour in blue tits

C. Steinmeyer · J. C. Mueller · B. Kempnaers

Received: 13 March 2008 / Accepted: 18 August 2008 / Published online: 16 September 2008
© The Author(s) 2008. This article is published with open access at Springerlink.com

Abstract The identification of functional polymorphisms in genes that underlie behavioural trait variation is a challenging but intriguing task in evolutionary biology. Given the wealth of genomic data and the increasing number of genotype–phenotype association studies in model organisms, one can ask whether and how this information can be used for non-model organisms. Here we describe two strategies to search for likely functional polymorphisms in candidate genes in a bird species that has been intensively studied by behavioural and population ecologists, the blue tit *Cyanistes caeruleus*. In the first approach we searched for repeating elements in coding regions of the genome using information about repeats in *Gallus gallus* genes. The rationale is that tandem-repeat elements have a high potential to be polymorphic and functional. The second strategy aimed to replicate reported genotype–phenotype association studies by extrapolating results from model organisms to our study species. Both strategies showed high success rates with respect to finding homologous gene regions and potentially informative genetic variants in the genes *AANAT*, *ADCYAP1*, *CKIε*, *CLOCK*, *CREB1*, *NPAS2* and *PERIOD2*.

Keywords *Cyanistes caeruleus* · Circadian · Clock · Polymorphism

Introduction

Genome-wide association studies are becoming an increasingly effective tool for identifying genetic factors contributing to complex traits (Amos 2007). This approach, however, is only applicable to genetic model species with available genome sequences and genome-wide polymorphism data. The majority of species including several ecological model species lack these data and it is unlikely that this situation will substantially change in the near future. Therefore the candidate gene approach is an appropriate choice when searching for functional or adaptively relevant polymorphisms in genetic non-model organisms (Fitzpatrick et al. 2005; Tabor et al. 2002). In this approach, functional genes identified from studies in genetic model organisms, such as *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus* or *Homo sapiens* serve as ‘candidate genes’ for similar phenotypic traits in other organisms. The structure of many genes and their cellular functions are highly conserved between evolutionary divergent animal taxa which makes this approach promising (Fitzpatrick et al. 2005). For example, information about the genetic basis of complex traits obtained from *Drosophila* has been used as a model for human traits and diseases (Mackay and Anholt 2006).

Once a gene is identified it is even more intriguing to find functional genetic polymorphisms in the non-model species that vary with the trait of interest. Various research strategies to detect functionally important genetic variation in natural populations have been proposed (Vasemagi and Primmer 2005), and several species-specific polymorphisms in candidate genes have been detected (e.g. Abzhanov et al. 2006; Fidler et al. 2007). The latter study replicated previously reported associations between a dopamine receptor D4 variant and human personality in a

Electronic supplementary material The online version of this article (doi:10.1007/s10709-008-9318-y) contains supplementary material, which is available to authorized users.

C. Steinmeyer · J. C. Mueller · B. Kempnaers (✉)
Max Planck Institute for Ornithology (Seewiesen), Postfach
1564, 82305 Starnberg, Seewiesen, Germany
e-mail: b.kempnaers@orn.mpg.de

wild bird species. Interestingly, the avian functional polymorphism—although different in type—was located in the same genomic region (exon) as the mammalian polymorphism (Fidler et al. 2007).

In this paper we describe two different strategies to find naturally occurring polymorphisms in candidate genes, which are likely to have functional consequences for circadian behavioural and physiological rhythms in birds. The circadian clock is perhaps the aspect of animal behaviour most fully characterised at the molecular level (Bell-Pedersen et al. 2005). This knowledge can be utilized in a candidate gene approach to analyse the influence of natural genetic variation on circadian behaviour. Our species of interest is the blue tit *Cyanistes caeruleus*, which is a common European passerine bird. Free-living populations of blue tits are studied by many research groups across Europe, mostly in the context of population or behavioural ecology. Johnsen et al. (2007) for example reported a polymorphic tandem repeat in the *CLOCK* gene, one of the core genes involved in generating endogenous rhythms. The authors reported a correlation between the *CLOCK* repeat length and latitude of different blue tit populations.

Strategy I: search for tandem repeats in exonic regions

Rationale and approach

The first strategy was to look for simple tandem repeat elements in exons of the candidate genes, either in protein coding regions or in untranslated regions (UTRs). It has been shown that the mutation rate of microsatellites is higher than that of non-repeat sequences (Jeffreys et al. 1988). These mutations are a consequence of different mechanisms such as unequal crossingover, gene conversion or replication slippage, that lead to a change in sequence length (reviewed in Nikitina and Nazarenko 2004). Several studies have shown that the number of repeats at mini-/microsatellite loci can influence different aspects of gene function. In particular, increasing length of trinucleotide repeats are associated with various inherited neurodegenerative disorders in humans (e.g. fragile X syndrome, Verkerk et al. 1991; Huntington disease, Mirkin 2007; Spinocerebellar ataxias, Orr and Zoghbi 2007 and Friedreich ataxia, Campuzano et al. 1996). Many of these disorders involve repeats which produce polyglutamine tracts in the amino acid sequence, i.e. they directly affect protein structure. Gene expression can also be altered by other mechanisms such as degradation of mRNA, decrease in protein production, an increase in DNA methylation resulting in the absence of gene expression, repressed transcription through increased nucleosome stability or

gene silencing (Choong et al. 1996; Imagawa et al. 1995; Pieretti et al. 1991; Wang et al. 1994). The relevance of tandem repeats in gene regulation was emphasized by a study showing that trinucleotide repeats in *Saccharomyces cerevisiae* are more frequent in open reading frames (ORFs) of genes that encode proteins involved in the regulation of transcription than in any other type of ORF (Young et al. 2000). Furthermore, dinucleotide repeats have the ability to form left-handed Z-DNA which plays a role in regulation of transcription due to altered DNA structure and protein binding affinity (Comings 1998).

In order to find repeating elements with a high probability of showing functional polymorphisms in clock genes, we first queried the National Center for Biotechnology Information (NCBI) database on the web (<http://www.ncbi.nlm.nih.gov/>) with the following key words: “biological rhythm* or biological timing or circadian rhythm* or central clock and Eukaryota”. Secondly, we queried the University of California Santa Cruz (UCSC) Genome browser (<http://www.genome.ucsc.edu/cgi-bin/hgGateway>) searching for “Simple tandem repeats within RefSeq exons of chicken”. The NCBI query resulted in a list of 206 genes that have been reported to have a function related to the endogenous clock. The UCSC query resulted in a list of 438 genes that are known to contain simple tandem repeats in exons of annotated genes of the chicken (*Gallus gallus*). The intersection between the two sets of genes contained the following five genes: *CLOCK* (circadian locomotor output cycles kaput protein), *NPAS2* (neuronal PAS domain protein 2), *ADCYAP1* (=PACAP, adenylate cyclase activating polypeptide 1), *CREB1* (cAMP responsive element binding protein 1, containing two distinct repeats), *CSNK1A1* (casein kinase 1 alpha 1) (see Table 1). *CSNK1A1* contains a repeat sequence with a period length of about 1,700 base-pairs. Therefore it is unlikely to show a polymorphism and consequently was not considered further.

We then BLASTed the chicken mRNA sequence of the exon comprising the repeat against the NCBI databases “nucleotide collection” and “Non-human, non-mouse ESTs” limiting the search to “Aves” and against the zebra finch (*Taeniopygia guttata*) WGS database of the NCBI trace archives (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>). The chicken sequence and an aligned sequence of a second bird species (in these cases sequences of zebra finch or wild turkey *Meleagris gallopavo*) were used to generate primer oligonucleotides flanking the requested exon. For designing forward and reverse primers for PCR amplification we used the program PrimaClade (<http://www.umsl.edu/services/kellogg/primaclade.html>), which is suited to design primers from multiple-species alignments. Primers were between 18 and 26 base-pairs in length and had one to three degenerated positions if necessary (Supplementary

Table). After amplifying the targeted sequence of the blue tit genome in a thermocycler (Supplementary Table) we ran the PCR products of each gene from 16 presumably unrelated adult individuals on a 10% polyacrylamide gel. If bands on the gel showed any inter-individual difference due to length variance of the amplified products we confirmed the presence of a polymorphism by running the fragments on an ABI 3130 sequencer. For this purpose we used fluorescently labelled forward primers in the PCR reactions.

Results

We amplified each of the requested exons of the four candidate genes in the blue tit genome (Table 1). Four of the five simple tandem repeats of interest were also found in the blue tit (Table 2). Furthermore, these repeats were polymorphic: in a sample of 148 presumably unrelated blue tits sampled in 2007 in our study population (Westerholz, 48°08' N 10°53' E, Southern Germany) the number of alleles varied between 5 and 7 (Table 2). Sequencing one carrier of each occurring allele confirmed that the detected

variation in length of the alleles is caused by different copy numbers of the repeating elements. Genotype proportions were in Hardy–Weinberg equilibrium (Table 2, all $P > 0.1$). We also tested whether the developed primers for the tandem repeats could be used in another songbird, the blackcap (*Sylvia atricapilla*). In a sample of 70 individuals from Southern France (43°31' N 4°43' E), we could directly genotype all four microsatellites. All markers were in Hardy–Weinberg equilibrium except for *NPAS2* which turned out to be monomorphic in this sample. Heterozygosity values ranged between 0.30 and 0.69.

Strategy II: search based on reported genotype–phenotype associations

Rationale and approach

The second strategy to find functionally important polymorphisms in genes involved in generating endogenous rhythms was based on reported genotype–phenotype associations in model organisms. First, we searched Web

Table 1 Genes involved in the endogenous clock that contain a repeat element in RefSeq exons of chicken

Gene name	Gene function	Repeat element	Average copy number	Position of repeat in gene of chicken
<i>ADCYAP1</i> (=PACAP)	Neurotransmitter	Dinucleotide	25	3' UTR
<i>CLOCK</i>	Transcription factor	Trinucleotide	22.3	Exon 20
<i>CREB1</i>	Transcription factor	Dinucleotide	24	3' UTR
		Dinucleotide	21.5	3'UTR
<i>CSNK1A1</i> ^a	Regulation of signal transduction	Repeat of 1,738 base-pairs	2.0	Exon 6 and adjacent intron
<i>NPAS2</i>	Transcription factor	Trinucleotide	8.3	Exon 20

^a Not considered further in our search strategy since the contained repeat element is unlikely to be polymorphic due to its period length

Table 2 Detected repeats in clock genes of blue tits

Gene name	Sequence of repeat 5' → 3'	No. of alleles	Allele size (bp)/frequency	$H_{Obs}^{b,c}$	$H_{Exp}^{b,c}$	χ^2c	P^c
<i>ADCYAP1</i>	(CT) ₅ TT(CT) _{12–16}	7	162/0.44, 164/0.42, 163 ^a /0.08, 166/0.03, 165 ^a /0.01, 158/0.01, 160/0.003	0.68	0.62	2.24	0.13
<i>CLOCK</i>	(CAA) ₂ (CAG) _{5–9} (CAA) _{1–2} CAGCAA	5	200/0.65, 203/0.18, 197/0.09, 206/0.07, 194/0.01	0.57	0.53	1.85	0.17
<i>CREB1</i>	(CA) _{6–7} CC(CA) ₃ (CC) _{0–1} ((CA) ₃ CC) _{0–1} (CA) _{2–9} TC/CC(CA) ₃	7	548/0.85, 546/0.05, 556/0.02, 540/0.02, 534/0.02, 550/0.01, 538/0.01	0.27	0.27	12.24	0.97
<i>NPAS2</i>	CAG(CAA) _{0–1} (CAG) _{4–7} (CAA) _{0–1} CAGCAACAG(CAA) ₂	5	178/0.39, 181/0.21, 184/0.19, 175/0.19, 172/0.02	0.75	0.73	2.75	0.84

Sequences and the number of alleles are based on a sample of $N = 148$ presumably unrelated individuals

^a Alleles with one base-pair difference in length are caused by at least one additional polymorphism next to the simple tandem repeat

^b Observed (Obs) and expected (Exp) heterozygosity

^c Calculated with Cervus 3.0 (<http://www.fieldgenetics.com/pages/home.jsp>)

of Science[®] (<http://apps.isiknowledge.com>) for articles reporting an association between polymorphisms in clock genes and a relevant phenotype. We used key words such as: “polymorphism AND sleep”, “circadian rhythm AND polymorphism”, “clock gene AND sleep”, “circadian rhythms AND gene”, “polymorphism AND clock” or in general “polymorphism AND bird AND association”.

In total, 24 studies tried to link a naturally occurring allelic variation in a clock gene with a phenotype that is influenced by the endogenous circadian clock (Table 3). Phenotypes investigated in humans were either extreme diurnal preferences classified as morningness or eveningness (Horne and Ostberg 1976), or a variety of sleeping patterns, including various sleep disorders (Table 3). In non-human animals, studies reported the period length of different behavioural rhythms. For further work, we selected those studies that reported a significant association between genotype and phenotype.

Second, we attempted to localize the position of the investigated polymorphism in the homologous bird gene. In the case of single nucleotide polymorphisms (SNPs) located in protein-coding regions, we proceeded as follows. First, we identified the position of the amino acid encoded by the polymorphic codon in the organism as reported in the original paper. We then aligned the amino acid sequence of the entire protein from the reported organism with the chicken genome in the UCSC browser. If we found a good alignment between the protein region of interest and a sequence in the chicken genome, we aligned the complete protein of chicken from the NCBI database entry to the chicken genome in the UCSC browser. In the resulting protein-DNA alignment the intron-exon structure of the chicken gene is shown, and we could determine the exon comprising the potentially polymorphic codon.

For SNPs located in UTRs of a gene and for the tandem repeat in exon 18 of *PERIOD 3* (see Table 3), the respective mRNA sequence of the model organism obtained from the NCBI database was directly aligned to the chicken genome. This was done by a NCBI “BLAST” (Altschul et al. 1997, Benson et al. 2008, <http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) and UCSC “BLAT” (Kent 2002, <http://www.genome.ucsc.edu/cgi-bin/hgBlat?command=start&org=Chicken&db=gallGal3&hgsid=104455251>) homology search. The best alignment was then tested for consensus in the region of the studied polymorphism.

To amplify the sequences of interest in the blue tit we designed PCR primers as described above for the tandem repeat sequences. The goal was to amplify the entire exon containing the homologous position of the polymorphism of interest. Degenerated primers were developed from the alignment between chicken and zebra finch, or, if no zebra finch sequence was available, between chicken and human,

Japanese quail (*Coturnix japonica*) or starling (*Sturnus vulgaris*) sequences (Supplementary Table). After PCR amplification of genomic DNA from blue tits (Supplementary Table), the PCR products of 10–14 presumably unrelated adult individuals were directly sequenced. Electropherograms were studied by manual inspection and sequences then aligned by using the programme Bioedit (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) and screened for sequence variations in either heterozygous or homozygous forms.

The ABI PRISM[®] SNaPshot[™] Multiplex Kit (Pati et al. 2004) was then used to genotype 149 presumably unrelated adults sampled in our population in 2007 for each detected exonic SNP.

Results

In total, we studied polymorphisms at 18 different sites (17 SNPs and one VNTR = variable number tandem repeats) in 10 genes for which a significant association to a behavioural trait was reported (Table 3). Fourteen of these published polymorphisms were discovered in human genes, the remaining four in *Drosophila*, mouse (*Mus musculus*) and Syrian hamster (*Mesocricetus auratus*). For seven of the SNPs we could identify the exact position in the homologous exon of four different chicken genes. At six sites the amino acid and the nucleotides in the mRNA coding for this amino acid were identical between chicken and the studied organism. At one SNP site (exon 4 of *AAVAT*; Table 3) the coded amino acid was not identical between chicken and the reported species, but we still found high similarity at the surrounding amino acids. We obtained a specific PCR product for the blue tit of these seven target regions. For the marker *CKIδ* no polymorphisms were detected in the sample of 10 blue tit sequences.

Overall, we discovered seven exonic SNPs in the amplified fragments of blue tit DNA in the genes *AAVAT* (2 SNPs), *PERIOD 2* (2 SNPs) and *CKIε* (3 SNPs). SNP sites were located between 1 and 79 base-pairs away from the position of the reported SNP in the model-organism. All detected SNPs in blue tits were silent (synonymous).

For six of these seven SNPs we genotyped 149 individuals and detected two alleles at each locus (Table 4). Genotyping of one SNP in *CKIε* failed so far. Genotype frequencies for the six SNPs were not significantly different from those expected under Hardy–Weinberg equilibrium (Table 4, all $P > 0.3$). We found significant linkage disequilibrium between the two SNPs in the gene *PERIOD2* and between the two SNPs in the gene *CKIε* ($D'_{(PERIOD2)} = 0.995$, $P_{(PERIOD2)} = 0.035$; $D'_{(CKIε)} = 0.998$, $P_{(CKIε)} < 0.0001$).

Table 3 Reported studies in different species that aimed to link a circadian phenotype with a certain polymorphism in one of the clock genes

Gene	Species	Region	Original name of polymorphism ^a	Amino acid change	Phenotype	Association	Publication
<i>CLOCK</i>	Human	Exon 17	A1982G	Yes	DSPS and N-24	No	Iwase et al. (2002)
		Exon 17	G1955A	Yes	DSPS and N-24	No	Iwase et al. (2002)
		3' UTR	T3111C	No	Morningness–eveningness	Yes	Mishima et al. (2005)
					Morningness–eveningness	Yes	Katzenberg et al. (1998)
					Morningness–eveningness, DSPS	No	Robilliard et al. (2002)
					DSPS and N-24	No	Iwase et al. (2002)
5' UTR	T257G	No	Morningness–eveningness, DSPS	No	Pedrazzoli et al. (2007)		
<i>PERIOD1</i>	Human	Exon 18	T2434C	No	Morningness–eveningness	Yes	Carpen et al. (2006)
		Exon 18	A2548G	No	Morningness–eveningness	No	Katzenberg et al. (1999)
<i>PERIOD2</i>	Human	Exon 17	A2106G	Yes	ASPS	Yes	Toh et al. (2001)
		Promotor	C-1228T	No	Morningness–eveningness	No	Carpen et al. (2005)
		5' UTR	C111G	No	Morningness–eveningness	Yes	Carpen et al. (2005)
		Exon 23	G3853A	Yes	morningness–eveningness	No	Carpen et al. (2005)
<i>PERIOD3</i>	Human	Exon 15, 17, 18, 20	T1940G, C2590G, T3110C, A3473A, del(3031–3084nt)	Yes	DSPS (haplotype analyses across all five loci)	Yes	Ebisawa et al. (2001)
		Exon 18	4-/5-repeat = del(3031–3084nt)		Morningness–eveningness	Yes	Jones et al. (2007)
		Exon 15	T1940G	Yes	Morningness–eveningness	Yes	Archer et al. (2003)
<i>AANAT</i> ^c	Human	Promotor	T-542G	No	Sleep pattern	No	Johansson et al. (2003)
		Promotor	G-263C	No	Sleep pattern	Yes	Wang et al. (2004)
					DSPS	No	Wang et al. (2004)
		Exon 4	C702T	No	DSPS	No	Hohjoh et al. (2003)
		Exon 4	C756T	No	DSPS	No	Hohjoh et al. (2003)
Exon 4	G619A	Yes	DSPS	yes	Hohjoh et al. (2003)		
<i>CKIδ</i> ^d	Human	Exon 2	T44A ^b	Yes	FASPS	Yes	Xu et al. (2005)
<i>CKIε</i> ^e	Human	Exon 9	G1223A	Yes	DSPS and N-24	Yes	Takano et al. (2004)
<i>MTNR1a</i> ^f	Human	Exon 1	C160T	Yes	DSPS and N-24	No	Ebisawa et al. (1999)
		Exon 2	C470T	Yes	DSPS and N-24	No	Ebisawa et al. (1999)

Table 3 continued

Gene	Species	Region	Original name of polymorphism ^a	Amino acid change	Phenotype	Association	Publication
<i>MTNR1b</i> ^g	Human	Exon 1	G71A	Yes	DSPS and N-24	No	Ebisawa et al. (2000)
		Exon 1	C196T	Yes	DSPS and N-24	No	Ebisawa et al. (2000)
<i>PRNP</i> ^h	Human	Exon 2	D178N ^b	Yes	Fatal familial insomnia	Yes	Tafti et al. (2005)
<i>TIMELESS</i>	Human	Exon 20	A2634G	Yes	Morningness–eveningness	No	Pedrazzoli et al. (2000)
<i>HLA DRI</i>	Human				DSPS	No	Hohjoh et al. (1999)
<i>doubletime</i>	<i>Drosophila</i>	Exon 3	dbt ^L	Yes	Period of behavioural rhythm	Yes	Kloss et al. (1998)
		Exon 3	dbt ^S	Yes	Period of behavioural rhythm	Yes	Kloss et al. (1998)
<i>Rab3a</i> ⁱ	Mouse	Exon 2	A3144G	Yes	Shortened circadian period	Yes	Kapfhamer et al. (2002)
<i>CKIε</i> ^e	Hamster	Pos. 532	R178C ^b	Yes	Free-running rhythm	Yes	Lowrey et al. (2000)

DSPS, Delayed sleep phase syndrome; (F)ASPS, (Familial) advanced sleep phase syndrome; N-24, Non-24-hour sleep-wake syndrome; SNP, Single nucleotide polymorphism; VNTR, Variable number of tandem repeats

^a Usually refers to nucleotide exchange and the nucleotide position

^b Amino acid change and its position in the protein

^c Arylalkylamine *N*-acetyltransferase

^d Casein kinase 1 delta

^e Casein kinase 1 epsilon

^f Melatonin 1a receptor

^g Melatonin 1b receptor

^h Prion protein

ⁱ *ras*-associated binding protein 3a

Table 4 Single nucleotide polymorphisms (SNPs) in three clock genes of blue tits

	Gene name		Nucleotide variants	Major allele frequency	$H_{\text{Obs}}^{\text{b,c}}$	$H_{\text{Exp}}^{\text{b,c}}$	χ^2_{c}	P^{c}
Major nucleotides are underlined. Sample size $N = 149$	<i>AANAT</i>	SNP1	<u>C</u> /T	0.91	0.18	0.17	1.48	0.36
		SNP2	G/ <u>A</u>	0.93	0.13	0.13	0.77	0.62
^a Information not yet available	<i>PERIOD2</i>	SNP1	G/ <u>A</u>	0.87	0.22	0.23	0.10	0.74
^b Observed (Obs) and Expected (Exp) heterozygosity		SNP2	<u>C</u> /T	0.91	0.17	0.17	0.05	1
^c Calculated with R (http://www.r-project.org) using the package “genetics”	<i>CKIε—tau</i>	SNP1	<u>C</u> /T	0.94	0.12	0.11	0.62	0.65
		SNP2	<u>C</u> /T	0.57	0.45	0.49	1.06	0.32
	<i>CKIε—dbt</i>	SNP3	T/ <u>G</u>	— _a	— _a	— _a	— _a	— _a

Discussion

We reported high success rates for two strategies to detect potentially functional polymorphisms in candidate genes of interest. The first approach was to search for tandem repeats in exonic regions of genes. All four target regions that contained tandem repeats in chicken genes could be amplified in the homologous blue tit genes. Backström et al. (2008) designed primers at conserved sites between

chicken and zebra finch, and found cross-species primer amplification success for blue tits of 83% ($N = 122$ markers tested). Difficulties referring to cross-species primer design turned out to be of minor importance in both strategies of our study. Only three out of eleven pairs of primers designed from chicken and another bird reference species did not amplify the homologous fragment in blue tits and this problem could be solved by designing new sets of primers. Furthermore, four of the five studied tandem

repeats could be found in the blue tit and they were polymorphic. Thus, gene-associated tandem repeats seem to be highly conserved among the class of Aves.

In contrast, it is estimated that only 13–25% of anonymous microsatellites developed for specific passerines co-amplify in other passerine species (Dawson et al. 2006; Hansson et al. 2005; Primmer et al. 1996). However, this relatively low cross-species amplification success may be due to variable primer binding sites, and not to the absence of the tandem repeat itself. In general, tandem repeats have a high potential to be functional and polymorphic (Contente et al. 2002; Iglesias et al. 2004; Nikitina and Nazarenko 2004). Therefore, this group of markers—if available—could be a first choice for testing genotype–phenotype associations.

Our second strategy was to focus on gene regions (particularly exons), for which positive associations with phenotypic traits had already been reported. Our analysis showed a high success rate in finding a polymorphism in the target region. In summary, we were able to detect 1–2 polymorphisms in four blue tit regions that showed high homology to the target regions of human, *Drosophila* and hamster. Although the detected SNPs in blue tit genes are silent, they might have functional consequences by mechanisms affecting mRNA structure (Duan et al. 2003; Kimchi-Sarfaty et al. 2007; Shen et al. 1999) or pre-mRNA splicing (Cartegni et al. 2002). In total we sequenced about 2,180 base-pairs of genomic blue tit DNA, of which 1,630 base-pairs were exonic and 550 base-pairs intronic. Thus SNPs in exons occurred on average every 233 base-pairs in the single blue tit population we studied. This is more frequent than the estimate reported for collared flycatchers *Ficedula albicollis* (on average one SNP in every 550 base-pairs of coding sequence; Backström et al. 2008), but it has to be considered that our calculation is based on a rather short DNA sequence of about 2.2 kb.

Both search strategies revealed only a small number of polymorphisms with the likelihood of being functional. This can now be tested by association with behavioural traits of interest. Thus, the advantage of this approach—in contrast to large-scale genotyping—is that it reduces the multiple testing problem and the costs. A prerequisite for these approaches is that the genes underlying phenotypic traits are already known. Many other strategies, for example QTL mapping (Flint and Mott 2001), candidate gene approach (Tabor et al. 2002) and targeted mutation (knock-out and knock-in technologies, Austin et al. 2004; Rago et al. 2007) have recently been developed to find the genetic basis of phenotypic traits. The approach described here allows us to investigate in a following study whether particular genetic polymorphisms are associated with variability in phenotypes, under the assumption that the function of a gene is conserved between different animal

taxa. The candidate-gene strategy is appropriate for the majority of organisms where no databases for genetic polymorphisms are available. A clear disadvantage of that approach in contrast to whole genome studies is that relevant polymorphisms might be missed.

Acknowledgements We thank Jacob Höglund and Gernot Segelbacher for organizing the ESF Workshop on “Ecological Genomics” and all participants at the workshop for discussion. We are grateful to Francisco Pulido for providing the blackcap samples, to Alexander Girg for technical assistance in the lab and to two anonymous reviewers for helpful comments on the manuscript. This work was supported by the Max Planck Society.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Abzhanov A, Kuo WP, Hartmann C et al (2006) The calmodulin pathway and evolution of elongated beak morphology in Darwin’s finches. *Nature* 442:563–567. doi:10.1038/nature04843
- Altschul SF, Madden TL, Schaffer AA et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. doi:10.1093/nar/25.17.3389
- Amos CI (2007) Successful design and conduct of genome-wide association studies. *Hum Mol Genet* 16:R220–R225. doi:10.1093/hmg/ddm161
- Archer SN, Robilliard DL, Skene DJ et al (2003) A length polymorphism in the circadian clock gene *Per3* is linked to delayed sleep phase syndrome and extreme diurnal preference. *Sleep* 26:413–415
- Austin CP, Battey JF, Bradley A et al (2004) The knockout mouse project. *Nat Genet* 36:921–924. doi:10.1038/ng0904-921
- Backström N, Fagerberg S, Ellegren H (2008) Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Mol Ecol* 17:964–980. doi:10.1111/j.1365-294X.2007.03551.x
- Bell-Pedersen D, Cassone VM, Earnest DJ et al (2005) Circadian rhythms from multiple oscillators: Lessons from diverse organisms. *Nat Rev Genet* 6:544–556. doi:10.1038/nrg1633
- Benson DA, Karsch-Mizrachi I, Lipman DJ et al (2008) GenBank. *Nucleic Acids Res* 36:D25–D30. doi:10.1093/nar/gkm929
- Campuzano V, Montermini L, Molto MD et al (1996) Friedreich’s ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271:1423–1427. doi:10.1126/science.271.5254.1423
- Carpen JD, Archer SN, Skene DJ et al (2005) A single-nucleotide polymorphism in the 5′-untranslated region of the *hPER2* gene is associated with diurnal preference. *J Sleep Res* 14:293–297. doi:10.1111/j.1365-2869.2005.00471.x
- Carpen JD, von Schantz M, Smits M et al (2006) A silent polymorphism in the *PER1* gene associates with extreme diurnal preference in humans. *J Hum Genet* 51:1122–1125. doi:10.1007/s10038-006-0060-y
- Cartegni L, Chew SL, Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3:285–298. doi:10.1038/nrg775

- Choong CS, Kempainen JA, Zhou ZX et al (1996) Reduced androgen receptor gene expression with first exon CAG repeat expansion. *Mol Endocrinol* 10:1527–1535. doi:10.1210/me.10.12.1527
- Comings DE (1998) Polygenic inheritance and micro/minisatellites. *Mol Psychiatry* 3:21–31. doi:10.1038/sj.mp4000289
- Contente A, Dittmer A, Koch MC et al (2002) A polymorphic microsatellite that mediates induction of PIG3 by p53. *Nat Genet* 30:315–320. doi:10.1038/ng836
- Dawson DA, Burke T, Hansson B et al (2006) A predicted microsatellite map of the passerine genome based on chicken-passerine sequence similarity. *Mol Ecol* 15:1299–1320. doi:10.1111/j.1365-294X.2006.02803.x
- Duan JB, Wainwright MS, Comeron JM et al (2003) Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum Mol Genet* 12:205–216. doi:10.1093/hmg/ddg055
- Ebisawa T, Kajimura N, Uchiyama M et al (1999) Allelic variants of human melatonin 1a receptor: Function and prevalence in subjects with circadian rhythm sleep disorders. *Biochem Biophys Res Commun* 262:832–837. doi:10.1006/bbrc.1999.1308
- Ebisawa T, Uchiyama M, Kajimura N et al (2000) Genetic polymorphisms of human melatonin 1b receptor gene in circadian rhythm sleep disorders and controls. *Neurosci Lett* 280:29–32. doi:10.1016/S0304-3940(99)00981-7
- Ebisawa T, Uchiyama M, Kajimura N et al (2001) Association of structural polymorphisms in the human period3 gene with delayed sleep phase syndrome. *EMBO Rep* 2:342–346. doi:10.1093/embo-reports/kve070
- Fidler AE, van Oers K, Drent PJ et al (2007) Drd4 gene polymorphisms are associated with personality variation in a passerine bird. *Proc R Soc Lond B Biol Sci* 274:1685–1691. doi:10.1098/rspb.2007.0337
- Fitzpatrick MJ, Ben-Shahar Y, Smid HM et al (2005) Candidate genes for behavioural ecology. *Trends Ecol Evol* 20:96–104. doi:10.1016/j.tree.2004.11.017
- Flint J, Mott R (2001) Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat Rev Genet* 2:437–445. doi:10.1038/35076585
- Hansson B, Akesson M, Slate J et al (2005) Linkage mapping reveals sex-dimorphic map distances in a passerine bird. *Proc R Soc Lond B Biol Sci* 272:2289–2298. doi:10.1098/rspb.2005.3228
- Hohjoh H, Takahashi Y, Hattai Y et al (1999) Possible association of human leucocyte antigen DR1 with delayed sleep phase syndrome. *Psychiatry Clin Neurosci* 53:527–529. doi:10.1046/j.1440-1819.1999.00574.x
- Hohjoh H, Takasu M, Shishikura K et al (2003) Significant association of the arylalkylamine *N*-acetyltransferase (AA-NAT) gene with delayed sleep phase syndrome. *Neurogenetics* 4:151–153
- Horne JA, Ostberg O (1976) A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Int J Chronobiol* 4:97–110
- Iglesias AR, Kindlund E, Tammi M et al (2004) Some microsatellites may act as novel polymorphic cis-regulatory elements through transcription factor binding. *Gene* 341:149–165. doi:10.1016/j.gene.2004.06.035
- Imagawa M, Ishikawa Y, Shimano H et al (1995) CTG triplet repeat in mouse growth-inhibitory Factor metalloproteinase-III gene promoter represses the transcriptional activity of the heterologous promoters. *J Biol Chem* 270:20898–20900. doi:10.1074/jbc.270.36.20898
- Iwase T, Kajimura N, Uchiyama M et al (2002) Mutation screening of the human Clock gene in circadian rhythm sleep disorders. *Psychiatry Res* 109:121–128. doi:10.1016/S0165-1781(02)00006-9
- Jeffreys AJ, Royle NJ, Wilson V et al (1988) Spontaneous Mutation-rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* 332:278–281. doi:10.1038/332278a0
- Johansson C, Willeit M, Smedh C et al (2003) Circadian clock-related polymorphisms in seasonal affective disorder and their relevance to diurnal preference. *Neuropsychopharmacology* 28:734–739. doi:10.1038/sj.npp.1300121
- Johnsen A, Fidler AE, Kuhn S et al (2007) Avian Clock gene polymorphism: evidence for a latitudinal cline in allele frequencies. *Mol Ecol* 16:4867–4880. doi:10.1111/j.1365-294X.2007.03552.x
- Jones KHS, Ellis J, Von Schantz M et al (2007) Age-related change in the association between a polymorphism in the PER3 gene and preferred timing of sleep and waking activities. *J Sleep Res* 16:12–16. doi:10.1111/j.1365-2869.2007.00561.x
- Kapfhamer D, Valladares O, Sun Y et al (2002) Mutations in Rab3a alter circadian period and homeostatic response to sleep loss in the mouse. *Nat Genet* 32:290–295. doi:10.1038/ng991
- Katzenberg D, Young T, Finn L et al (1998) A CLOCK polymorphism associated with human diurnal preference. *Sleep* 21:569–576
- Katzenberg D, Young T, Lin L et al (1999) A human period gene (HPER1) polymorphism is not associated with diurnal preference in normal adults. *Psychiatr Genet* 9:107–109. doi:10.1097/00041444-199906000-00011
- Kent WJ (2002) BLAT - The BLAST-like alignment tool. *Genome Res* 12:656–664. doi:10.1101/gr.229202. Article published online before March 2002
- Kimchi-Sarfaty C, Oh JM, Kim IW et al (2007) A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science* 315:525–528. doi:10.1126/science.1135308
- Kloss B, Price JL, Saez L et al (1998) The Drosophila clock gene double-time encodes a protein closely related to human casein kinase I epsilon. *Cell* 94:97–107. doi:10.1016/S0092-8674(00)81225-8
- Lowrey PL, Shimomura K, Antoch MP et al (2000) Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. *Science* 288:483–491. doi:10.1126/science.288.5465.483
- Mackay TEC, Anholt RRR (2006) Of flies and man: Drosophila as a model for human complex traits. *Annu Rev Genomics Hum Genet* 7:339–367. doi:10.1146/annurev.genom.7.080505.115758
- Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* 447:932–940. doi:10.1038/nature05977
- Mishima K, Tozawa T, Satoh K et al (2005) The 311T/C polymorphism of hClock is associated with evening preference and delayed sleep timing in a Japanese population sample. *Am J Med Genet B Neuropsychiatr Genet* 133B:101–104. doi:10.1002/ajmg.b.30110
- Nikitina TV, Nazarenko SA (2004) Human microsatellites: mutation and evolution. *Russ J Genet* 40:1065–1079. doi:10.1023/B:RUGE.0000044750.21421.65
- Orr HT, Zoghbi HY (2007) Trinucleotide repeat disorders. *Annu Rev Neurosci* 30:575–621. doi:10.1146/annurev.neuro.29.051605.113042
- Pati N, Schowinsky V, Kokanovic O et al (2004) A comparison between SNaPshot, pyrosequencing, and biplex invader SNP genotyping methods: accuracy, cost, and throughput. *J Biochem Biophys Methods* 60:1–12. doi:10.1016/j.jbbm.2003.11.005
- Pedrazzoli M, Ling L, Finn L et al (2000) A polymorphism in the human timeless gene is not associated with diurnal preferences in normal adults. *Sleep Res (Online)* 3:73–76
- Pedrazzoli M, Louzada FM, Pereira DS et al (2007) Clock polymorphisms and circadian rhythms phenotypes in a sample of the Brazilian population. *Chronobiol Int* 24:1–8. doi:10.1080/07420520601139789
- Pieretti M, Zhang FP, Fu YH et al (1991) Absence of expression of the Fmr-1 gene in Fragile-X syndrome. *Cell* 66:817–822. doi:10.1016/0092-8674(91)90125-I

- Primmer CR, Moller AP, Ellegren H (1996) A wide-range survey of cross-species microsatellite amplification in birds. *Mol Ecol* 5:365–378. doi:[10.1111/j.1365-294X.1996.tb00327.x](https://doi.org/10.1111/j.1365-294X.1996.tb00327.x)
- Rago C, Vogelstein B, Bunz F (2007) Genetic knockouts and knockins in human somatic cells. *Nat Protocols* 2:2734–2746. doi:[10.1038/nprot.2007.408](https://doi.org/10.1038/nprot.2007.408)
- Robilliard DL, Archer SN, Arendt J et al (2002) The 3111 Clock gene polymorphism is not associated with sleep and circadian rhythmicity in phenotypically characterized human subjects. *J Sleep Res* 11:305–312. doi:[10.1046/j.1365-2869.2002.00320.x](https://doi.org/10.1046/j.1365-2869.2002.00320.x)
- Shen LX, Basilion JP, Stanton VP (1999) Single-nucleotide polymorphisms can cause different structural folds of mRNA. *Proc Natl Acad Sci USA* 96:7871–7876. doi:[10.1073/pnas.96.14.7871](https://doi.org/10.1073/pnas.96.14.7871)
- Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 3:391–A396. doi:[10.1038/nrg796](https://doi.org/10.1038/nrg796)
- Tafti M, Maret S, Dauvilliers Y (2005) Genes for normal sleep and sleep disorders. *Ann Med* 37:580–589. doi:[10.1080/07853890500372047](https://doi.org/10.1080/07853890500372047)
- Takano A, Uchiyama M, Kajimura N et al (2004) A missense variation in human casein kinase I epsilon gene that induces functional alteration and shows an inverse association with circadian rhythm sleep disorders. *Neuropsychopharmacology* 29:1901–1909. doi:[10.1038/sj.npp.1300503](https://doi.org/10.1038/sj.npp.1300503)
- Toh KL, Jones CR, He Y et al (2001) An hPer2 phosphorylation site mutation in familiar advanced sleep phase syndrome. *Science* 291:1040–1043. doi:[10.1126/science.1057499](https://doi.org/10.1126/science.1057499)
- Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Mol Ecol* 14:3623–3642. doi:[10.1111/j.1365-294X.2005.02690.x](https://doi.org/10.1111/j.1365-294X.2005.02690.x)
- Verkerk A, Pieretti M, Sutcliffe JS et al (1991) Identification of a gene (Fmr-1) containing a Cgg repeat coincident with a breakpoint cluster region exhibiting length variation in Fragile-X syndrome. *Cell* 65:905–914. doi:[10.1016/0092-8674\(91\)90397-H](https://doi.org/10.1016/0092-8674(91)90397-H)
- Wang YH, Amirhaeri S, Kang S et al (1994) Preferential nucleosome assembly at DNA triplet repeats from the myotonic-dystrophy gene. *Science* 265:669–671. doi:[10.1126/science.8036515](https://doi.org/10.1126/science.8036515)
- Wang GY, Lee CGL, Lee EJD (2004) Genetic variability of arylalkylamine-*N*-acetyl-transferase (AA-NAT) gene and human sleep/wake pattern. *Chronobiol Int* 21:229–237. doi:[10.1081/CBI-120037822](https://doi.org/10.1081/CBI-120037822)
- Xu Y, Padiath QS, Shapiro RE et al (2005) Functional consequences of a CKI delta mutation causing familial advanced sleep phase syndrome. *Nature* 434:640–644. doi:[10.1038/nature03453](https://doi.org/10.1038/nature03453)
- Young ET, Sloan JS, Van Riper K (2000) Trinucleotide repeats are clustered in regulatory genes in *Saccharomyces cerevisiae*. *Genetics* 154:1053–1068