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

C. Steinmeyer · J. C. Mueller · B. Kempenaers

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
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 behaviour 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 is 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 Saccharomyes 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 Eu karyota”. 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), CREBI (cAMP responsive element binding protein 1, containing two distinct repeats), CSNKIA1 (casein kinase 1 alpha 1) (see Table 1). CSNKIA1 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 |
| ADCYAP1 (=PACAP) | Neurotransmitter | Dinucleotide | 25 | 3’ UTR |
| CLOCK | Transcription factor | Trinucleotide | 22.3 | Exon 20 |
| CREB1 | Transcription factor | Dinucleotide | 24 | 3’ UTR |
| CSNK1A1 | Regulation of signal transduction | Repeat of 1,738 base-pairs | 2.0 | Exon 6 and adjacent intron |
| NPAS2 | 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 | \(H_{\text{Obs}}\) | \(H_{\text{Exp}}\) | \(\chi^2\) | \(P\) |
| ADCYAP1 (CT)\textsubscript{9}TT(CT)\textsubscript{12–16} | 7 | 162/0.44, 164/0.42, 163/0.08, 166/0.03, 165/0.01, 158/0.01, 160/0.03 | 0.68 | 0.62 | 2.24 | 0.13 |
| CLOCK (CAA)\textsubscript{3}(CAG)\textsubscript{5–9}(CAA)\textsubscript{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 |
| CREB1 (CA)\textsubscript{6–7}CC(CA)\textsubscript{3}(CA)\textsubscript{0–1}(CA)\textsubscript{3}CAGCAA | 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 |
| NPAS2 CAG(CAA)\textsubscript{1–2}(CAG)\textsubscript{5–9}(CAA)\textsubscript{0–1} | 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=galGal3&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 AANAT; 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 AANAT (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 (\( D'_{(PERIOD2)} = 0.995, p_{(PERIOD2)} = 0.035; D'_{(CKIε)} = 0.998, p_{(CKIε)} < 0.0001 \)).
| Gene   | Species | Region       | Original name of polymorphism\(a\) | Amino acid change | Phenotype                        | Association | Publication                  |
|--------|---------|--------------|-----------------------------------|------------------|----------------------------------|-------------|-----------------------------|
| **CLOCK** | 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)    |
|        |         |              |                                   |                  | DSPS and N-24                    | No          | Robilliard et al. (2002)    |
|        |         |              |                                   |                  | DSPS and N-24                    | No          | Iwase et al. (2002)         |
|        |         |              |                                   |                  | Morningness–eveningness, DSPS    | No          | Pedrazzoli et al. (2007)    |
|        |         | 5' UTR       | T257G                             | No               | Morningness–eveningness, DSPS    | No          | Pedrazzoli et al. (2007)    |
| **PERIOD1** | Human   | Exon 18      | T2434C                            | No               | Morningness–eveningness          | Yes         | Carpen et al. (2006)        |
|        |         | Exon 18      | A2548G                            | No               | Morningness–eveningness          | No          | Katzenberg et al. (1999)    |
| **PERIOD2** | Human   | Exon 17, Promotor | A2106G                          | Yes              | ASPS                             | Yes         | Toh et al. (2001)           |
|        |         |              | 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)        |
| **PERIOD3** | 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, DSPS | Yes | Morningness–eveningness, DSPS | Yes | Jones et al. (2007) |
|        |         | Exon 15, T1940G | Morningness–eveningness, DSPS | Yes | Morningness–eveningness, DSPS | Yes | Archer et al. (2003) |
| **AANAT** | Human   | Promotor     | T-542G                            | No               | Sleep pattern                    | No          | Wang et al. (2004)          |
|        |         | Promoter     | G-263C                            | No               | Sleep pattern                    | Yes         | 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)        |
| **CK1δ** | Human   | Exon 2       | T44A\(^{b}\)                       | Yes              | FASPS                             | Yes         | Xu et al. (2005)            |
| **CK1ε** | Human   | Exon 9       | G1223A                            | Yes              | DSPS and N-24                    | Yes         | Takano et al. (2004)        |
| **MTNR1A** | 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)       |
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

Table 3 continued

| Gene | Species | Region | Original name of polymorphism | Amino acid change | Phenotype | Association | Publication |
|------|---------|--------|-------------------------------|------------------|-----------|-------------|-------------|
| MTNR1b | 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) |
| PRNP | Human | Exon 2 | D178N \( b \) | Yes | Fatal familial insomnia | Yes | Tafti et al. (2005) |
| TIMELESS | Human | Exon 20 | A2634G | Yes | Morningness–eveningness | No | Pedrazzoli et al. (2000) |
| HLA DRI | Human | | | | DSPS | No | Hohjoh et al. (1999) |
| doubletime | Drosophila | Exon 3 | dbt \( a \) | Yes | Period of behavioural rhythm | Yes | Kloss et al. (1998) |
| | | Exon 3 | dbt \( a \) | Yes | Period of behavioural rhythm | Yes | Kloss et al. (1998) |
| Rab3a | Mouse | Exon 2 | A3144G | Yes | Shortened circadian period | Yes | Kapfhamer et al. (2002) |
| CKIε | 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

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

| Gene name | Nucleotide variants | Major allele frequency | \( H_{\text{obs}} \) | \( H_{\text{Exp}} \) | \( \chi^2 \) | \( P \) |
|-----------|--------------------|-----------------------|----------------|-------------|----------------|--------|
| AANAT SNP1 | C/T | 0.91 | 0.18 | 0.17 | 1.48 | 0.36 |
| SNP2 | G/A | 0.93 | 0.13 | 0.13 | 0.77 | 0.62 |
| PERIOD2 SNP1 | G/A | 0.87 | 0.22 | 0.23 | 0.10 | 0.74 |
| SNP2 | C/T | 0.91 | 0.17 | 0.17 | 0.05 | 1 |
| CKIε—tau SNP1 | C/T | 0.94 | 0.12 | 0.11 | 0.62 | 0.65 |
| SNP2 | C/T | 0.57 | 0.45 | 0.49 | 1.06 | 0.32 |
| CKIε—dbt SNP3 | T/G | — | — | — | — | — |

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 Girt 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:10.1111/j.1365-294X.2007.03551.x

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.1365-294X.2005.00471.x

Carpen JD, van 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
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

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

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

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

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

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

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

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

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

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

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

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

Young ET, Sloan JS, Van Riper K (2000) Trinucleotide repeats are clustered in regulatory genes in Saccharomyces cerevisiae. Genetics 154:1053–1068