# Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
BIOLOGIA EVOLUZIONISTICA ED AMBIENTALE

CICLO XXVIII

COORDINATORE Prof. Guido Barbujani

# Evolution of clock genes in human populations

Settore Scientifico Disciplinare BIO/18

**Dottorando**
Dott. Dall'Ara Irene

**Tutore**
Prof. Bertolucci Cristiano

Anni 2013 / 2015

# Evolution of Clock Genes in Human populations

**TABLE OF CONTENTS:**


1. INTRODUCTION: p. 1
   - 1.1 Biological rhythms: p. 1
     - 1.1.1 Origin and evolution p. 1
     - 1.1.2 Molecular architecture of the circadian clock p. 5
     - 1.1.3 Central and peripheral oscillators p. 9
     - 1.1.4 Circadian phenotypes: the chronotypes p. 13
     - 1.1.5 Polymorphisms associated with chronotypes p. 17
     - 1.1.6 Abnormal circadian rhythms: FASPS and DSPS p. 19
   - 1.2 Elements of population genetics: p. 23
     - 1.2.1 The origins of Population Genetics p. 23
     - 1.2.2 Genetic variability and Hardy-Weinberg equilibrium p. 24
     - 1.2.3 Genetic diversity, admixture and population structure p. 26
     - 1.2.4 Mutation generates genetic diversity p. 28
     - 1.2.5 Natural Selection p. 28

2. AIM OF THE THESIS: p. 31

3. MATERIALS AND METHODS: p. 33
   - 3.1 Genetic data: the 1000Genomes Project p. 33
   - 3.2 Environmental data p. 36
   - 3.3 Analysis of population structure: p. 38
     - 3.3.1 Principal Component Analysis p. 38
     - 3.3.2 ADMXITURE p. 39
   - 3.4 Testing for the effect of natural selection p. 40
   - 3.5 Effect of natural selection in the phenotype-associated dataset p. 42

4. RESULTS: p. 44
   - 4.1 Analysis of population structure: p. 44
     - 4.1.1 Principal Component Analysis p. 44
     - 4.1.2 ADMXITURE p. 45
   - 4.2 Testing for the effects of natural selection p. 47
   - 4.3 Effect of natural selection in the phenotype-associated dataset p. 50

5. DISCUSSION p. 51

6. FUTURE PERSPECTIVES p. 56

7. LITERATURE CITED p. 58

# 1. INTRODUCTION

## *1.1 Biological rhythms*

### 1.1.1 Origin and evolution

Earth's revolution and rotation lead to periodic environmental changes giving rise to many geophysical cycles, including daily (24 hours), seasonal (365 days), lunar (29.5 days) and tidal (12.8 hours) cycles (Ciarleglio et al., 2008; Foster and Roenneberg, 2008; Mongrain and Cermakian, 2009). These environmental modifications led to quantitative and qualitative oscillations in metabolic, behavioural and physiological functions. Organisms anticipate upcoming daily changes to have the maximum benefit from limited natural resources. For instance, during the daylight many plants release oxygen as a photosynthesis product and birds of prey hunt. Conversely, in darkness conditions, cyanobacteria begin nitrogen fixation, nocturnal rodents start to forage and filamentous fungi begin their daily production of spores (Burh and Takahashi, 2013). However, the changes in the physical environment are not the only factors determining daily changes in behaviour and physiology. There is, indeed, a timekeeping system within the organisms, named the "biological clock". The examination of these periodic (cyclic) phenomena in living organisms and the study of mechanisms underlying the biological timekeeper system constitute the basic features of chronobiology (Frederick, 2015). In its broadest sense, it encompasses all the research areas on biological timing, ranging from the high-frequency cycles that characterize hormone secretion, to annual ones involved in the reproductive cycle in same species. Among them, rhythms with a period of about 24 hours, which give rise to daily cycles, are controlled by the so-called circadian clock (Mongrain and Cermakian, 2009).

The term "*circadian*" is derived from the Latin words "*circa diem*", meaning "about a day" (Bollinger and Schibler, 2014). In synchrony with the light-dark cycle, plankton travel up and down the water column, plants open and close their blossoms, and animals sleep and rest (Roenneberg and Merrow, 2005). From cyanobacteria to humans, a circadian system restricting behaviour to specific intervals in the solar day is present. The circadian clock has been found to be ubiquitous, and almost every cell in the body in a broad array of biological processes and organisms contains a circadian clock, with similar properties and even similar phase-response curve to light (Vitaterna et al., 2001; Mohawa et al., 2012). Its conservation across many diverse biological taxa, plants as well as animals, testifies to the importance of the system, which allow organisms to maximize energy efficiency while minimizing environmental risks variant within distinct niche-specific ranges over the solar cycle (Bedont and Blackshaw, 2015).

The first indication that daily rhythms were endogenously programmed, and not just a passive consequence to natural light or temperature cycles, date back the 18$^{th}$ century when the French astronomer Jean Jacques d'Ortous de Mairan reported daily leaf movements of the *Mimosa pudica* (Angiospermae, Fabaceae). He put a plant having a cycle of opening and closing its leaves every twenty-four hourssynchronized with the sunlight, into a dark closet. He observed the opening and closing of the leaves continued on a 24 hours' schedules, as if the light of the sun were present (De Mairan, 1729) demonstrating the independence of the biological rhythms to the natural light-dark cycle in living organisms. In this way he illustrated one of the most important features of circadian rhythms: the self-sustained nature (Vitaterna et al., 2001). More recently, this characteristic has been demonstrated also by Ko and coworkers (Ko et al., 2010), who have found that there is a network capable of generating oscillations in the circadian domain in the complete absence of cell autonomous oscillatory potential (Mohawk et al., 2012). The intrinsic oscillation of the circadian pacemaker when it is not influenced by environmental time cues, such as light and temperature, give rise to the so-called "free-running" (Sack et al., 2000).

In almost all living organism, the circadian clock drives cycles of behaviour and physiology with a period of approximately, but not exactly, 24 hours (Saini et al., 2015). For example, in humans and mice the average free-running period length ($\tau$) of the circadian clock is about 24.3 hours and 23.7 hours, respectively (Saini et al., 2015). Although investigations in constant conditions are important to understand mechanisms that generate the self-sustained oscillation, understanding of how biological clocks function under entrained conditions is the key for understanding their full biological significance (Roenneberg et al., 2005). As Pittendrigh (1960) demonstrated, the deviation from a 24-hour cycle actually provides a means for the internal timekeeping system to be continuously aligned by and aligned to the environment. Indeed, similar to physical oscillators, the circadian system actively synchronizes the temporal sequence of biological functions with the environmental cycles through signals called "*Zeitgebers*" (literally, "time giver" in German). Although the light-dark cycle clearly is the major *Zeitgeber* for all organisms, other factors, such as social interactions, availability of resources, activity or exercise, and even temperature, can also modulate the circadian cycle. This continuous adjustment, named *entrainment*, results in greater precision in controlling the timing, or phase, of the expressed rhythms. Accordingly, if a shift in external cues occurs (i.e. following travel across time zones), circadian rhythm is regularly adjusted, with a process known as phase shift (Sack et al., 2000). As an example, twice a year we get a reminder of the importance of our internal circadian biological clocks. Daylight savings: in October we fall back just an hour, and yet we wake up an hour early on Monday anyway and think meals are late-but only for a day, until our clocks are reset (Dunlap, 1999). The oscillatory behaviour of the system ensures that

entrainment is not passive or driven and therefore allows for great plasticity and adaptive potential (Roenneberg et al., 2003b). Initially, it was unclear whether entrainment was achieved by modulating the rate of cycling (i.e., whether the cycle was shortened or lengthened until it was aligned to the new cues and then reverted to its original length) or whether entrainment was achieved by discrete resetting events. In 1960 Pittendrigh discovered that the organism's response to light (i.e., whether a cycle advances, is delayed, or remains unchanged) differs depending on the phase in the cycle at which it is presented. Thus, exposure to light during the early part of the individual's "normal" dark period generally results in a phase delay, whereas exposure to light during the late part of the individual's normal dark period generally results in a phase advance (Figure 1).

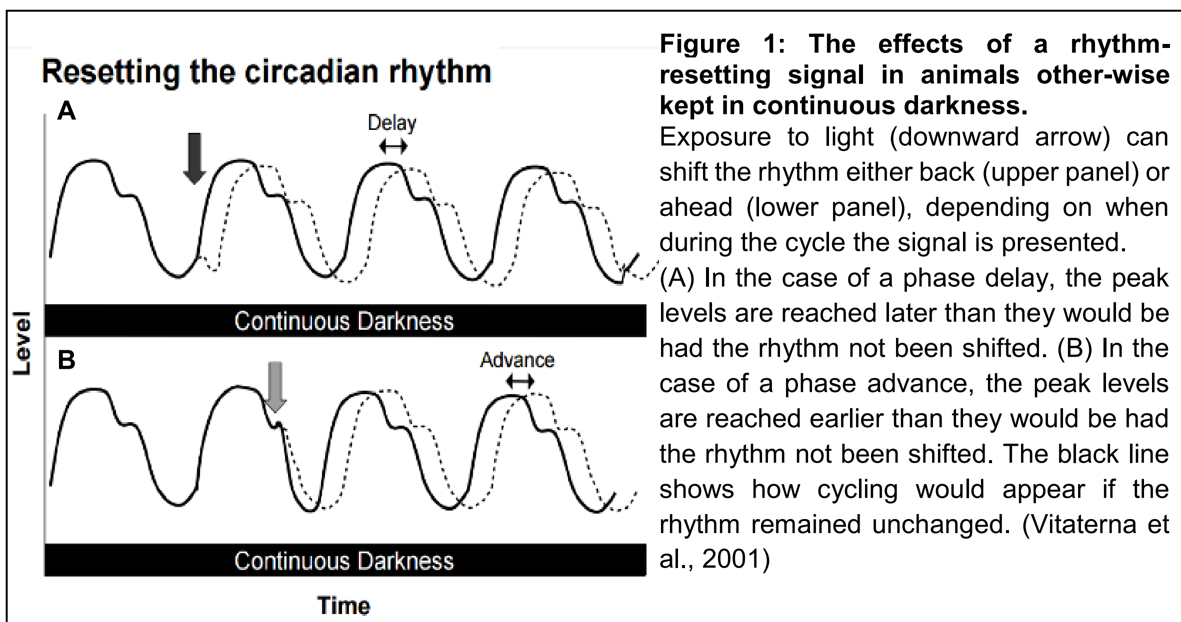



**Figure 1: The effects of a rhythm-resetting signal in animals other-wise kept in continuous darkness.**
Exposure to light (downward arrow) can shift the rhythm either back (upper panel) or ahead (lower panel), depending on when during the cycle the signal is presented.
(A) In the case of a phase delay, the peak levels are reached later than they would be had the rhythm not been shifted. (B) In the case of a phase advance, the peak levels are reached earlier than they would be had the rhythm not been shifted. The black line shows how cycling would appear if the rhythm remained unchanged. (Vitaterna et al., 2001)

The characteristics of entrainment by light (e.g., its phase relationship to the zeitgebers) depend on the fact that light shifts the circadian phase by different amounts and in different directions depending on when the circadian clock is exposed to light. Using single light exposures in otherwise constant conditions, so-called phase response curves (PRCs; Figure 2) can be established that indicate for each circadian time how much the clock is advanced (by definition, a positive value) or delayed (by definition, a negative value) (Roenneberg et al., 2003a). That is, such curves can predict the manner in which an organism will entrain not only to shifts in the light-dark cycles but also to unusual light cycles, such as non-24-hour cycles or different light:dark ratios. Depending on intensity and duration of the light signal, circadian systems respond more or less strongly, leading to steeper or flatter PRCs.

The phase of entrainment depends on how much and in what direction the free-running period deviates from 24 h, that is, how much the daily light signal has to advance or delay the clock. Thus, the phase of entrainment depends both on the strength of the *Zeitgeber*

and on the individual's free-running period. The existence of a phase-response curve also implies that entrainment is achieved by discrete resetting events rather than changes in the rate of cycling (Vitaterna et al., 2001).

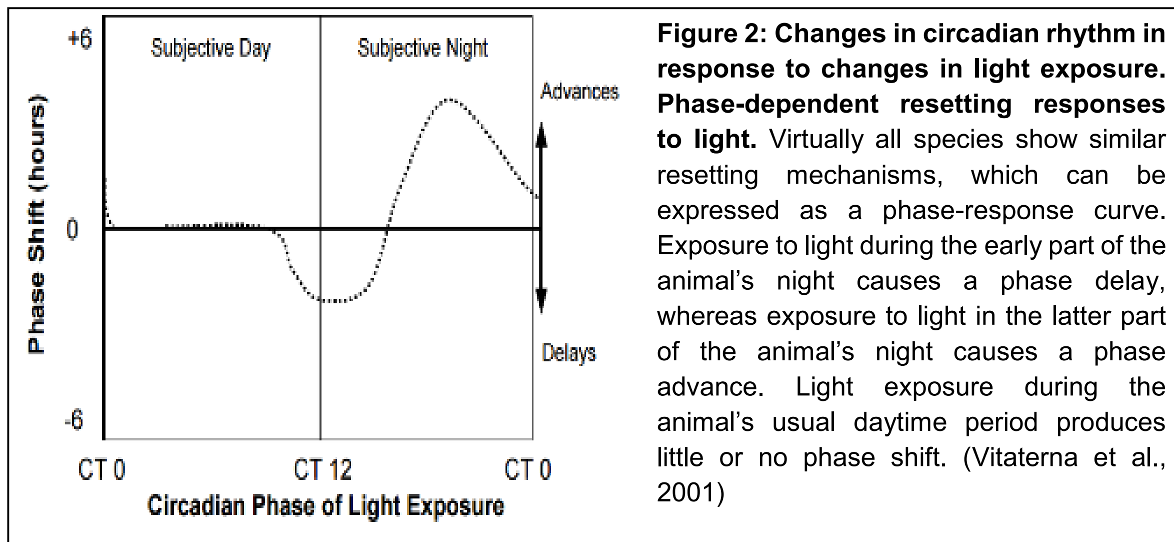



**Figure 2: Changes in circadian rhythm in response to changes in light exposure. Phase-dependent resetting responses to light.** Virtually all species show similar resetting mechanisms, which can be expressed as a phase-response curve. Exposure to light during the early part of the animal's night causes a phase delay, whereas exposure to light in the latter part of the animal's night causes a phase advance. Light exposure during the animal's usual daytime period produces little or no phase shift. (Vitaterna et al., 2001)

One of the hallmarks of circadian rhythms in all organisms, from cyanobacteria to mammals, is that they are temperature compensated — daily rhythms remain constant as temperature increases or decreases across a physiologically viable range (Pittendrigh, 1993). Thus, a change in temperature can affect the phase of a cycle without substantially altering the rate of cycling. This means that the cycle may start at an earlier or later-than-normal time but still have the same length (Vitaterna et al., 2001). This allows the circadian clock to provide an accurate measure of the passage of time regardless of ambient temperature. This does not, however, mean that the clock is insensitive to temperature since oscillator period length does alter, but not at the same rate as would be expected by most biochemical reactions (Edwards et al., 2005). It represents a buffer against inappropriate signals and, finally, this riddle has provided a clue to the nature of the internal clock— that is, the fact that circadian rhythms have a genetic basis. Such a program of gene expression would be more resistant to temperature alteration than, for example, a simple biochemical reaction (Vitaterna et al., 2001). Mammals do, however, experience circadian rhythms in core body temperature with a fluctuation of 1–4°C that are regulated by the SCN (Lowrey and Takahashi, 2011). This clock property can be expressed quantitatively as the $Q_{10}$, or temperature coefficient— the change in the rate of a biological rhythm or biochemical reaction as a result of increasing the ambient temperature by 10°C. The $Q_{10}$ for most biochemical reactions is 2 to 3, yet circadian rhythms usually have a $Q_{10}$ between 0.8–1.2 (Sweeney and Hastings, 1960).

### 1.1.2 Molecular architecture of the circadian clock

At the molecular level, the circadian clock involves interlocked positive and negative transcriptional/translational feedback loops between clock genes and their protein products, together with a multi-level post-translational regulation of key clock components (Mohawk et al., 2012).

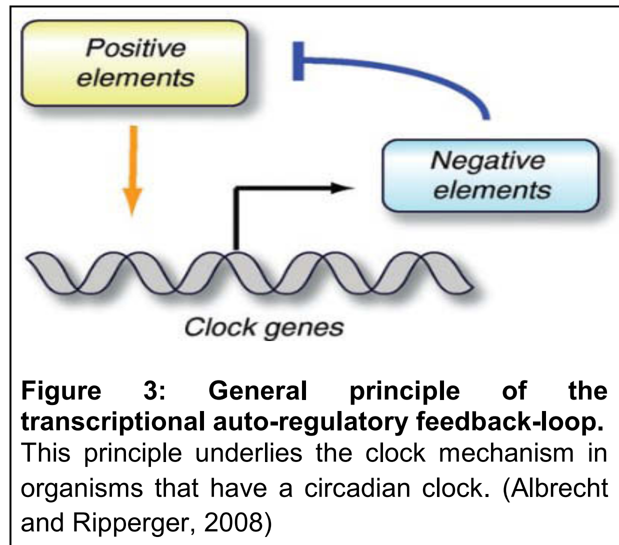



**Figure 3: General principle of the transcriptional auto-regulatory feedback-loop.** This principle underlies the clock mechanism in organisms that have a circadian clock. (Albrecht and Ripperger, 2008)

The simplest version of the hypothesis describes an auto-regulating negative feedback loop based on transcription and translation: a clock gene is transcribed, its RNA is translated into protein, the protein enters the nucleus, eventually inhibiting its own transcription until degradation of both RNA and protein relieves the inhibition and the cycle restarts (Figure 3). Over the years, this basic model has rapidly evolved with the discovery of other clock components in flies, as well as in fungi, plants, cyanobacteria and many representatives throughout the animal kingdom. Clock-gene families that were created by gene duplications were found in animals and plants. New mechanisms (for example, positive feedback loops) and additional, coupled feedback loops (for example, see Young and Kay, 2001) were postulated (Roenneberg and Merrow, 2005).

Concisely, ribonucleic acid (RNA) molecules undergo post-transcriptional processing in the nucleus. The messenger RNAs (mRNA) are exported to the cytoplasm where ribosomal translation forms proteins. Their stability is affected by kinase-mediated protein phosphorylation that takes place in the cytoplasm. After a time delay, sufficient protein molecules accumulate in the cytoplasm and form heterodimers that are imported into the nucleus, where they regulate gene expression through closure of positive and negative feedback loops (Voinescu, 2009). More precisely, among the two interconnected feedback loops in clock genes expression, the positive is characterized by a heterodimeric transcriptional complex driving the expression of genes which negatively feedback to inhibit their own transcription (Figure 4). This auto-regulated feedback loop, located at the intracellular level, takes ~24 hours to complete a cycle (Sato et al., 2006).

However, new insights have come from work on mammals, specifically from studying the daily activity profiles of clock mutant mice; from more detailed recordings of clock gene expression under different experimental conditions and in different tissues; and from the discovery and analysis of a growing number of additional clock genes. These new results

are moving the model paradigm away from a simple negative feedback loop to a molecular network (Roenneberg and Merrow, 2003).

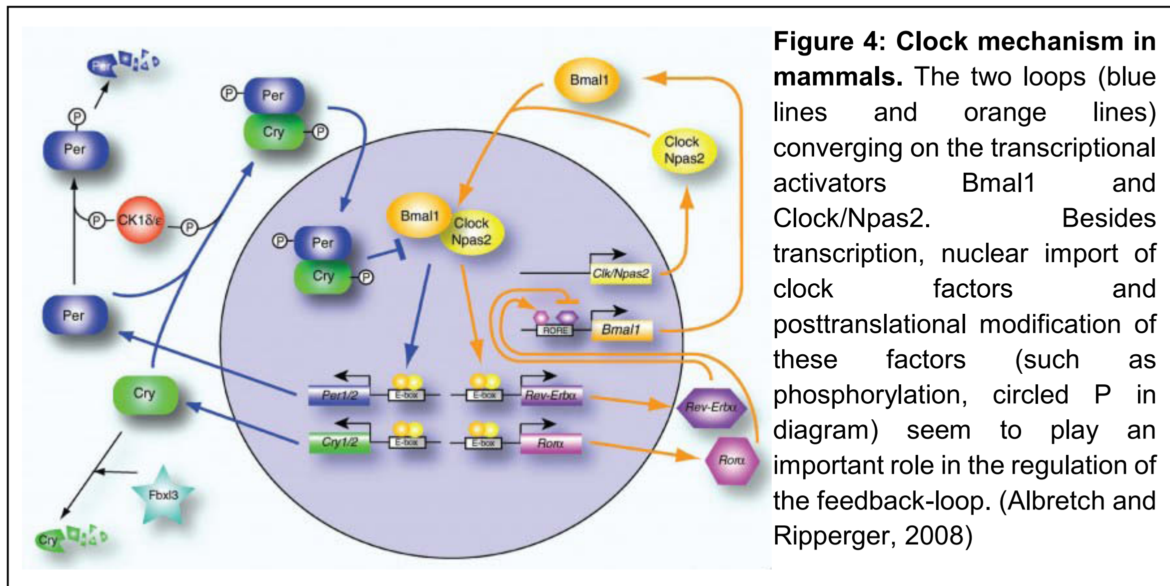



**Figure 4: Clock mechanism in mammals.** The two loops (blue lines and orange lines) converging on the transcriptional activators Bmal1 and Clock/Npas2. Besides transcription, nuclear import of clock factors and posttranslational modification of these factors (such as phosphorylation, circled P in diagram) seem to play an important role in the regulation of the feedback-loop. (Albretch and Ripperger, 2008)

The molecular basis of the circadian clock has firstly been clarified in *D. melanogaster*, and the discovered basic feedback regulation of the circadian clock is highly conserved among insects (Numata et al., 2015). The clock involves at least three interlocked transcriptional and translational feedback loops, the major of which consist of four core clock genes: *Period* (*Per*), *Timless* (*Tim*), *Clock* (*Clk*), and *Cycle* (*cyc*). In this loop, *Clk* and *Cyc* proteins, CLK and CYC respectively, form a heterodimer in order to activate transcription of *Per*, *Tim*, and clock-controlled genes. Therefore, they are referred to as positive elements. On the other hand, the products of *Per* and *Tim* proteins also form a heterodimer and function as negative elements by suppressing the activity of CLK/ CYC.

The change in the level of expression of core clock genes and their proteins occurs with a period of approximately 24 hours, generating circadian rhythms id daily events (Figure 5; Numata et al., 2015).

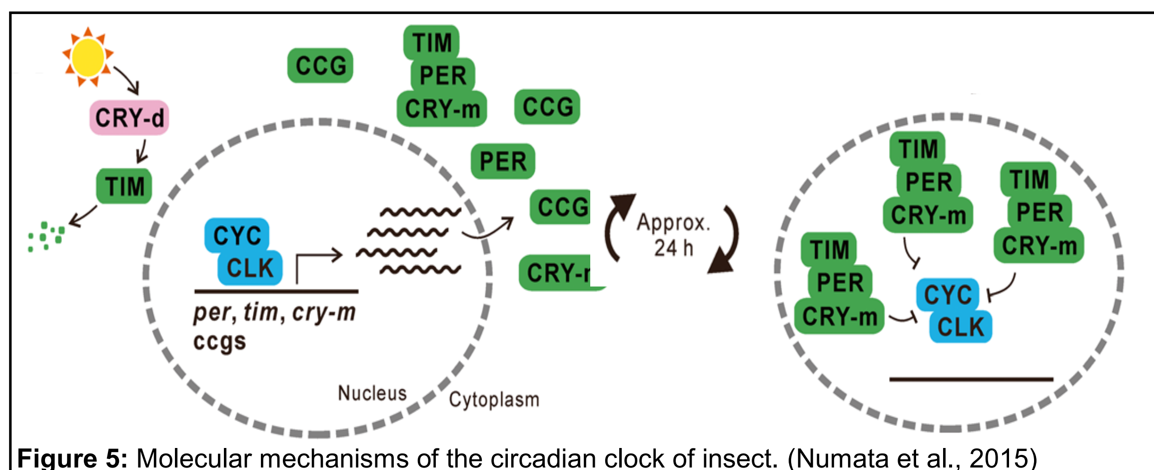



**Figure 5:** Molecular mechanisms of the circadian clock of insect. (Numata et al., 2015)

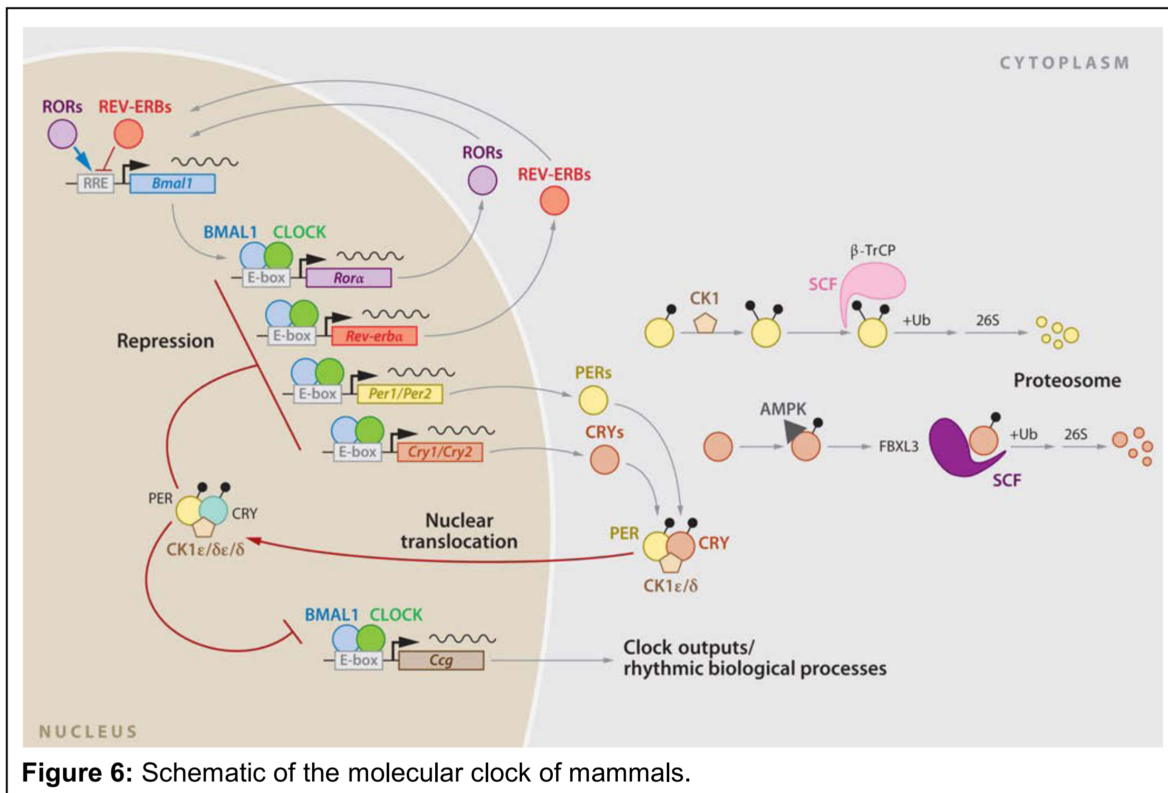



**Figure 6:** Schematic of the molecular clock of mammals.

In mammals, as shown in Figure 6, the positive loop consists of the cytoplasmic protein BMAL1 (Brain and muscle Arnt-like protein-1) that heterodimerizes with the basic HLH-PAS proteins CLOCK (Circadian Locomotor Output Cycles Kaput) or its paralogue NPAS2 (Neuronal PAS domain protein 2) and, the heterodimer binds the E-boxes (5'-CACGTG-3'), and E'-boxes (5'-CACGTT-3') elements in the promoters of target genes, namely the core clock gene families *Period* (*Per1*, *Per2*, and *Per3*) and *Cryptochrome* (*Cry1* and *Cry2*) (Lowrey and Takahashi, 2011). The members of the negative limb, in particular the PERs, act as the state variable in the mechanism (Edery et al. 1994). Briefly, this means that the levels of these proteins determine the phase of the clock (Buhr and Takahashi, 2013). The resulting PER and CRY proteins dimerize and translocate back to the nucleus. Here, they repress their own transcription by inhibiting the activity of the CLOCK:BMAL1 complexes (Figure 6) giving rise to the negative loop and allowing the cycle to repeat from a level of low transcriptional activity.

The chromatin remodelling necessary for this cyclic transcriptional activity is realized through a combination of clock-specific and ubiquitous histone modifying proteins and can be observed in the rhythmic acetylation/deacetylation of histones (H3 and H4) at multiple clock target genes (Sahar and Sassone-Corsi 2013). The histone acetyl transferase (HAT) domain located on the CLOCK protein is necessary for the rescue of rhythms in Clock-mutant fibroblasts (Doi et al. 2006). The CLOCK/BMAL1 complex also recruits the methyltransferase MLL1 to cyclically methylated histone H3 and HDAC inhibitor JARID1a to further facilitate transcriptional activation (DiTacchio et al. 2011). Deacetylation takes place, in part, due to recruitment by PER1 of the SIN3- HDAC (SIN3-histone deacetylase)

complex to CLOCK/BMAL1-bound DNA (Duong et al. 2011). Interestingly, the rhythmic deacetylation of histone H3 at the promoters of circadian genes is regulated by the deacetylase SIRT1, which is sensitive to NAD+ levels (Nakahata et al. 2008). This is remarkable considering that the NAD+ to NADH ratio has been shown to regulate CLOCK/ BMAL1's ability to bind DNA *in vitro* (Rutter et al. 2001). Thus, cellular metabolism may prove to play an important role in regulating the transcriptional state, and therefore the phase, of the clock (Marcheva et al. 2013). Degradation of the negative limb proteins PER and CRY is required to terminate the repression phase and restart a new cycle of transcription. Therefore, the stability/degradation rate of the PER and CRY proteins is a key element to setting the period of the clock.

Moreover, the CLOCK/BMAL1 dimers also initiate the transcription of a second feedback loop which acts in coordination with the loop described above. This involves the E-box-mediated transcription of the orphan nuclear-receptor genes Rev-Erbα/β and RORα/β (Buhr and Takahashi, 2013). Their gene products compete to binding ROREs elements (retinoic acid-related orphan receptor response elements) present within the *Bmal1* promoter, in order to regulate *Bmal1* expression (Lowrey and Takahashi, 2004). More precisely, ROR proteins initiate *Bmal1* transcription and REV-ERB proteins inhibit it (Buhr and Takahashi, 2013). Consequently, the circadian oscillation of *Bmal1* is both positively and negatively regulated by RORs and REV-ERBs. Casein kinase 1 epsilon and Casein kinase 1 delta (CK1ε and CK1δ) are critical factors that regulate the core circadian protein turnover in mammals.

Post-translational modifications of the core clock components play a crucial role in generating the delay necessary to establish the ~24 hours rhythms of the mammalian circadian clock (Lowrey et al., 2011). Among them, phosphorylation of clock proteins was the first process observed in the mammalian molecular clock. This modification, mediated by Casein kinase 1δ/ε, regulates PER subcellular localization and its ability to repress CLOCK/BMAL1-mediated transcription. For instance, polymorphisms in CK1ε and CK1δ are related to an alteration of the length of the circadian period in both mice and humans (Xu et al., 2005; Lee et al., 2009). Furthermore, phosphorylation promotes its ubiquitin-degradation (Ohsaki et al., 2008), one of the major pathway for protein degradation in cell. The system requires the attachment of multiple ubiquitin molecules to lysine residues on the target protein. The resulting polyubiquinated proteins are directed to the proteasome (26S) to been degraded in small peptides (Nandi et al., 2006). Another extensive post-translational modification that BMAL1 protein undergone is the sumoylation, a process that covalently attach small ubiquitin-like modifier (SUMO) proteins at lysine residues of target proteins to modify their function (Wilkinson and Henley, 2010). Sumoylation has been

shown to affect nuclear/cytosolic localization of proteins, progression through the cell cycle, protein stability, and transcriptional regulation (Gareau and Lima, 2010). Furthermore, this modification localizes BMAL1 to the promyelocytic leukemia nuclear body, potentiates CLOCK/BMAL1 transactivation of clock-controlled genes, and promotes BMAL1 ubiquitin-dependent proteasomal degradation (Lee et al., 2008).

The circadian clock consists also of an additional group of genes, and their proteins, not directly involved in the core clock mechanism, but linked with it. Among these, there are both clock-controlled (CCG) and clock-related (CRG) genes. An example is the mammalian TIM (TIMLESS) protein that, instead, are at the heart of molecular pathways important for chromosome integrity, efficient cell growth and/or development. It interacts with the core clock complex CRY:PER, but its functional link to the circadian clock is still under debate (Engelen et al., 2013). At the suprachiasmatic level, there are two important helix-loop-helix transcription factors: DEC1 and DEC2. Both of them suppress CLOCK:BMAL1-induced activation of *Per* gene (Honma et al., 2002). In addition, AANAT (arylalkylamine N-acetyltransferase) protein acts as the penultimate enzyme in the synthesis of melatonin. This is an endocrine hormone, secreted by pineal gland and, in animals, synchronizes several biological functions to the circadian rhythms. In Humans, it is important for regulating the sleep-wake cycle (Ciarleglio et al., 2008) and the onset of melatonin secretion under dim light conditions (DLMO) is the single most accurate marker for assessing the circadian pacemaker (Pandi-Perumal et al., 2007). Among all the clock components, some are expressed rhythmically, including the negative feedback loops and some activators, while other are expressed constitutively, such as the kinases and other activators (Roenneberg and Merrow, 2005).

### 1.1.3 Central and peripheral oscillators

The study of mechanisms underlying the biological timekeeper system began roughly 50 years ago and ranks Colin Pittendrigh and Jürgen Aschoff as firsts to investigate experimentally the circadian clocks in fruit fly and humans, respectively (Vitaterna et al., 2001). In one of Aschoff early studies (Aschoff, 1960), human subjects lived in a converted former bomb shelter underground without time cues or direct social interactions and could select their own sleep-wake schedules. The subjects exhibited an approximately 25-h rest-activity cycle as well as circadian oscillations of other physiological variables, suggesting the presence of an endogenous circadian clock. However, the mechanism for this endogenous clock remained largely unclear until so years later when a culmination of data from animal studies suggested a transcriptional-transitional feedback loop with a circadian period length (Hastings et al., 2008).

Although in the 1960s and 1970s there were debates on whether biological rhythms have an endogenous nature (Brown et al., 1970), currently the existence of internal clocks that control the circadian system on a cellular level is undisputed (Bedont and Blackshaw, 2015; Numata et al, 2015).

Many experiments have been done, both with animal and with plants model organisms, in order to better understanding the molecular bases of the circadian clock. In particular, pioneering work on *Drosophila melanogaster* has provided a basis for understanding how the temporal sequence of daily events is controlled in mammals. This has been possible thank to the fact that the underlying principle and the hierarchical structure of the circadian system have been highly maintained during evolution (Figure 7). In fact, in all circadian systems identified to date, regardless of phylogenetic origin, three major components are present: 1) a light input pathway to a self-sustained master circadian pacemaker, 2) the circadian pacemaker itself, and 3) and output pathways by which the circadian pacemaker regulates overt rhythms in biochemistry, physiology, and behaviour throughout the organism (Lowrey et al., 2004)

Indeed, different centres that control circadian physiology have been localised in the nervous system of many animals, from cockroaches to mammals. To cite some instances, in Lepidoptera and Diptera the central clock is localised in the mid- brain whereas in mammals the centre acting as the master pacemaker for the generation of circadian behavioural rhythms is the suprachiasmatic nucleus (SCN) of the hypothalamus (Mohawk et al., 2012; Numata et al., 2015).

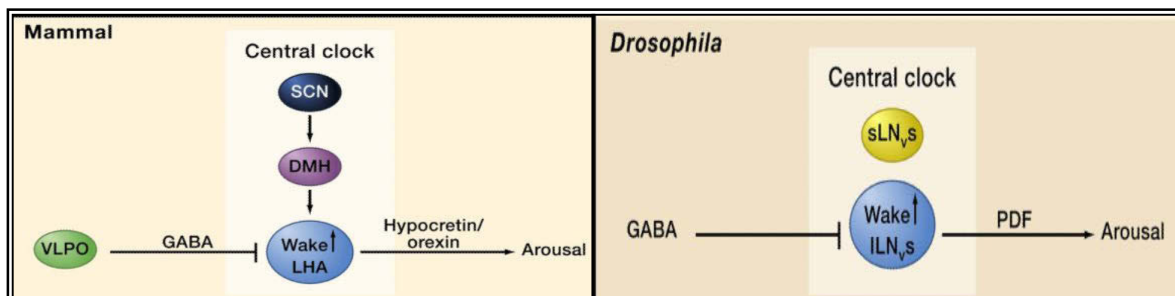



**Figure 7**: **Example of a conserved mechanism underlying sleep**
In both *Drosophila* and mammals, an arousal promoting peptide (PDF and Hypocretin, respectively) is secreted by cells within, or in the vicinity of, the central clock network.
In mammals, Hypocretin-producing neurons in the lateral hypothalamus receive circadian inputs from the central clock in the suprachiasmatic nucleus (SCN) via the Dorsomedial Hypothalamus (DMH). They are inhibited by GABAergic inputs from the ventrolateral preoptic (VLPO) area.
In *Drosophila*, the large ventral lateral neurons (lLNvs) are part of the clock network although they are not required for free-running circadian rhythms. Instead they mediate light-driven arousal, at least in part through the release of PDF. As in mammals, GABAergic inputs to these neurons promote sleep (Sehgal and Mingot, 2012)

The SCN is a paired neuronal structure located in the anteroventral hypothalamus, on either side of the third ventricle, just above the optic chiasm. In the rat and humans, the

suprachiasmatic nucleus contains of ~20,000 and ~200,000 neurons respectively (Bollinger and Schibler, 2014). These neurones are communicating with each other by synaptic and paracrine mechanisms or gap junctions, so as to maintain phase coherence between their clocks. $Ca^{2+}$ and cyclic adenosine monophosphate (cAMP) signalling (O'Neill and Reddy, 2012) as well as signalling by vasointestinal peptide (VIP) through the VPAC2 receptor (Harmar, 2003) are well-established paracrine mechanisms through which SCN cells couple their oscillators and keep them in synchrony (Bollinger and Schibler, 2014). In each unilateral SCN neurons are divided in two anatomic subdivisions: a ventral "core" region which abuts the optic chiasm and receives retinal input, and a dorsal "shell" region which partially envelops and receives input from the core (Welsh et al., 2010). The SCN receives direct photic input from the retina through a photoreceptor cell type known as intrinsically photoreceptive retinal ganglion cells (ipRGCs) (Do and Yau, 2010). These cells receive the non-visual inputs from rods and cones and express the photopigment *Opn4* (Melanopsin) which renders them intrinsically photosensitive to short wavelength irradiation (Mohawk et al., 2012). In humans, this centre resides a couple of centimetres behind the bridge of the nose, in a pair of nuclei above the crossing of the optic nerves. Each of these suprachiasmatic nuclei (SCN) of the hypothalamus is only about the size of a grain of rice, but, through their coupling, they appear to be responsible for organising endogenous daily programmes throughout the body (Mohawk et al., 2012).

To fulfil its central role in controlling circadian rhythms, the mammalian SCN uses a complex, interdependent hierarchy of signalling molecules to integrate information form assorted afferents and the multiple, robust clocks located in its subdivisions (Mohawk et al., 2012). In addition to the core transcriptional feedback loop in the SCN, circadian oscillations in the expression of clock genes can be observed throughout all the cells of the body (virtually, 3-4 x $10^{13}$ circadian oscillators; Bollinger and Schibler, 2014) although the phase of the rhythms is tissue-specific (Yoo et al., 2004).Hundreds and thousands of genes are expressed with a circadian rhythm in various tissue that possess the same functional and self-sustained nature as core circadian clocks (Gaspar and Brown, 2015).

All together, these rhythms make up the Peripheral clocks, which seems to be beneficial for anticipating external stimuli and thus contributes to the maintenance of organ homeostasis (Figure 8). The cellular circadian oscillators in all cells types share the same molecular mechanism (Bollinger and Schibler, 2014). The peripheral oscillators, however, differ from the SCN pacemaker in several ways. First, the phase of the peripheral clock oscillation is delayed by 3–9 hours relative to the SCN (Zylka et al., 1998). Second, ex vivo cultures of peripheral tissues from transgenic rats in which the mouse Per1 promoter drives a luciferase reporter gene reveal circadian rhythms in luminescence that damp out after 2–7 days (Yamazaki et al., 2000).

Given the hierarchical architecture of the circadian system, the SCN is believed to synchronize peripheral timekeepers through a wide variety of mechanisms, Loss of synchronization between the central and peripheral clocks also augments disease progression (Takeda and Maemura, 2015). As a result, the circadian clock exerts broad-ranging control over many biological processes, including many aspects of metabolism such as xenobiotic detoxification (Gachon et al., 2006), glucose homeostasis (Lamia et al., 2008; Marcheva et al., 2010; So et al., 2009; Turek et al., 2005), and lipogenesis (Gachon et al., 2011; Le Martelot et al., 2009).

In insects, peripheral circadian clocks reside in prothoracic glands, testes, Malpighian tubules, and epidermal cell (Numata et al., 2015). Experiments with mice and rats revealed that feeding cycles are dominant Zeitgebers (timing cues) for the oscillators operative in most peripheral organs, including liver, pancreas, kidney, heart, lung and skeletal muscles (Stokkan et al., 2001; Figure 8). Thus when feeding cycles are switched from night-time feeding – which is normal in nocturnal rodents – to daytime feeding, the phase of circadian gene expression in peripheral organs becomes gradually inverted after a week or two. Since feeding time has little impact on the phase of the SCN, peripheral oscillators become uncoupled from the SCN under these conditions.

Even mammals' circadian rhythms are characterized by peripheral oscillators, spread in almost all the body tissue and organs. They are synchronized through many ways, including metabolites (e.g. nicotinamide adenine dinucleotides), cytokines, neuronal signals from the

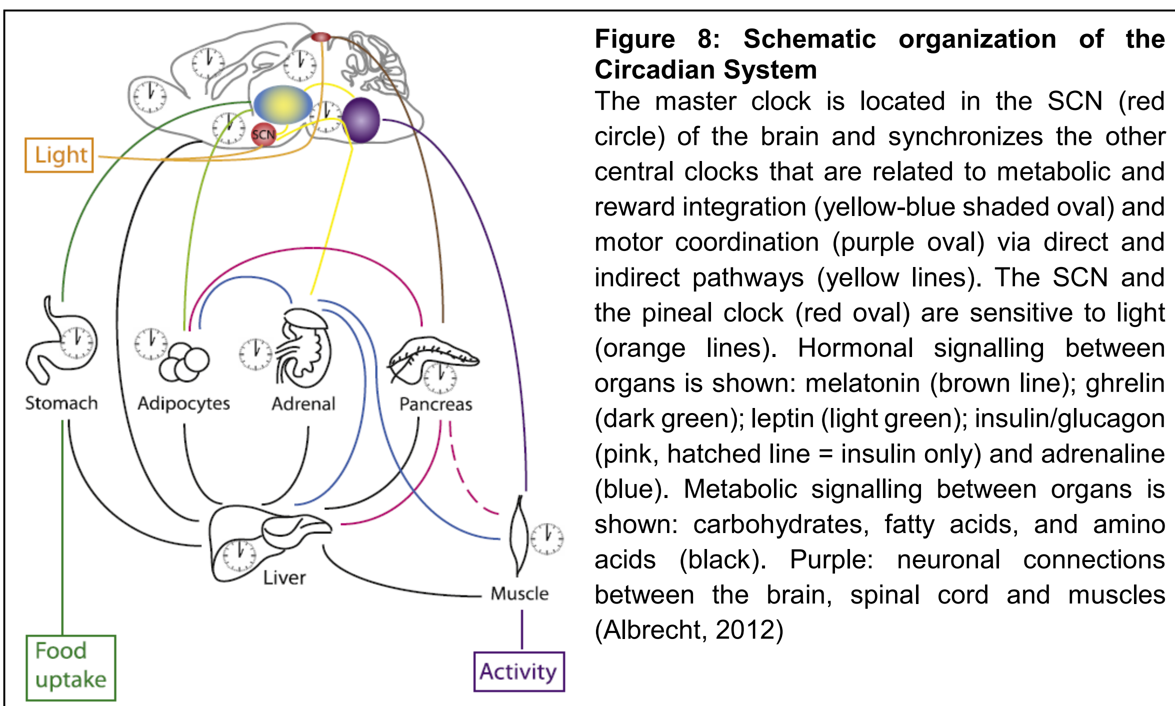



**Figure 8: Schematic organization of the Circadian System**
The master clock is located in the SCN (red circle) of the brain and synchronizes the other central clocks that are related to metabolic and reward integration (yellow-blue shaded oval) and motor coordination (purple oval) via direct and indirect pathways (yellow lines). The SCN and the pineal clock (red oval) are sensitive to light (orange lines). Hormonal signalling between organs is shown: melatonin (brown line); ghrelin (dark green); leptin (light green); insulin/glucagon (pink, hatched line = insulin only) and adrenaline (blue). Metabolic signalling between organs is shown: carbohydrates, fatty acids, and amino acids (black). Purple: neuronal connections between the brain, spinal cord and muscles (Albrecht, 2012)

peripheral nervous system and hormones. Among the last category, cortisol in humans and corticosterone in rodents can activate the glucocorticoid receptor which binds to

glucocorticoid-responsive elements within promoter and enhancer sequences of *Per1* and *Per2* genes. In general, glucocorticoids antagonise nutrient-dependent signals when the phase of feeding rhythms is in conflict with the phase of the SCN (Bollinger and Schibler, 2014).

Another, perhaps unexpected, zeitgebers for peripheral clocks is the body temperature. The rhythmic expression of temperature-dependent genes and its regulation through the SCN was firstly discovered from Kornmann and collaborators in a study on the genome-wide analysis of mRNAs whose cyclic expression in the mouse liver is driven by systemic cues rather than local hepatocyte clocks (Kornmann et al., 2007). Peripheral oscillators, including fibroblasts, liver, kidney, and lung, are exquisitely sensitive to temperature changes. They can be strongly reset by low amplitude temperature pulses that mimic the range of circadian variation; indeed, simulated mouse body temperature rhythms oscillating smoothly between 35°C and 38°C are sufficient to synchronise circadian oscillators of cultured cells (Buhr et al., 2010). Moreover, simulated human body temperature cycles fluctuating between 36°C and 37°C are able to synchronise the clocks of at least a fraction of cultured cells (Saini et al., 2012).

Nevertheless, there are specie-specific differences relative to the components of this mechanism and the role of each one. Indeed, cryptochrome is a light receptor in plants, but it functions as a transcriptional regulator in mammals and serves both functions in *D. melanogaster* (Roenneberg and Merrow, 2005).

### 1.1.4 Circadian phenotypes: the chronotypes

It was at the beginning of the 20th century when the botanist Wilhelm Johannsen proposed two new words, “genotype” and “phenotype” (Groth and Weiss, 2006). Respectively, these terms refer to the genetic constitution of an individual and to his/her visible aspect, as descriptors of the process of inheritance (genome), and to the process of development (phenome) including the physical or biochemical expressions of the genotype, as previously hypothesized by Mendel in the 19$^{th}$ century. When referred to the circadian clock, the external expression of the genotype is defined “chronotypes”.

The chronotypes (also referred to as circadian type, diurnal preference or diurnal variation) represents the tendency of each individual to be active early or late in the day (Saini et al., 2015). Human circadian conditions are classified according to the altered phase alignment between the internal clock and the external day-night cycle (Zhang et al., 2011). This phase of entrainment, is affected by both an individual’s free-running period and the duration of Zeitgebers themselves. As an example, the duration of light per day, called photoperiod, can drastically change with seasons; but also the amount of exposure to daily light changes

between farmer and office workers (Roenneberg and Merrow, 2005). Chronotype is heritable, as estimated by twin and family studies (12-42%), but its genetic basis has not yet been well defined (Lane et al., 2016).

Chronotypes have been assessed mainly by questionnaires, among them the Horne-Östberg (Horne and Östberg, 1976) and the Munich Chronotypes (Figure 9; Roenneberg et al., 2003a), designed to associate individuals to tendencies that were termed "morningness" or "eveningness", on the basis of a score. The value of Horne-Östberg questionnaire tests, for examples, range between 16-30, defined as "night type", and 70-86, defined as "morning type" (Toh et al., 2001). The questions used are mostly subjective, relating sleep and activity times to a personal "feeling best rhythm" (Horne and Östberg, 1976) or to the habits of others (e.g., "I get up later than most people") (Smith et al., 2002); in same cases the subjects are asked how they would behave under hypothetical situations (e.g., "Approximately what time would you get up if you were entirely free to plan your day?") (Terman and White, 2001). Chronotypes identified by the Home-Östberg questionnaire correlate with body temperature rhythms (Home and Östberg, 1976) and circadian period (Duffy et al., 2001) in selected, but not all, populations.

The circadian organization of the sleep/wake cycles and physiology appears during the first several months of life and it change during the lifetime. The childhood is characterized by earliness, late shifting to lateness, reaching its maximum around the age of 20. After that, there is a gradual return to early preferences. However, until around the age of 50, there are sex differences with women exhibiting generally earlier chronotypes than men (Saini et al., 2015). More recently, Hu and co-workers (Hu et al., 2016) results showed the prevalence in morning type is about one third in under 30-years-old individuals compared to over 60-years-old (24% and 63%, respectively). Although humans are by nature diurnal, there is a wide spectrum of sleep–wake time preferences in the general population, giving rise to three major circadian types: morning, intermediate and evening (Figure 10). All of them present differences in behaviour, physiology and personality (Voinescu, 2009). About 40% of the adult population is classified in one of the two extreme groups, while 60% is characterized by an intermediated chronotype (Adan et al., 2009).

The so-called "Morning type" (Figure 10 B), or "lark", wakes up spontaneously early in the morning, is more active in the first part of the day and tends to go to bed early in the evening. From a physiological point of view, the Morning type shows a steeper rise in body temperature and reaches her/his peak approximately 1 to 3 hours earlier than evening types (Bailey and Heitkemper, 2001). Circadian type appear to be strongly associated with the melatonin acrophase, with morning types having a more rapid decline in melatonin levels after the peak than evening types (Gibertini et al., 1999).

## Munich ChronoType Questionnaire (MCTQ)

Please enter your age, gender, etc.. This information is important for our evaluations

**Age:** _________ **female** **male** **Height** ______ **Weight** _______

**On work days ...**

I have to get up at... _________o'clock
I need... _________min to wake up
I regularly wake up... before the alarm with the alarm
From... _________o'clock I am fully awake
At around... _________o'clock, I have an energy dip
On nights before workdays, I go to bed at _________o'clock...
... and it then takes me... _________min to fall asleep

If I get the chance, I like to take a siesta/nap ...
correct I then sleep for..._________ min
not correct I would feel terrible afterwards

**On free days (please only judge normal free days, i.e., without parties etc.) ...**

My dream would be to sleep until... _________o'clock
I normally wake up at... _________o'clock
If I wake up at around the normal (workday) alarm time, I try to get back to sleep...
correct not correct
If I get back to sleep, I sleep for another... _________min
I need ... _________min to wake up
From... _________o'clock I am fully awake
At around... _________o'clock, I have an energy dip
On nights before free days, I go to bed at... _________o'clock...
... and it then takes me... _________min to fall asleep

If I get the chance, I like to take a siesta/nap ...
correct I then sleep for..._________ min
not correct I would feel terrible afterwards

once I am in bed, I would like to read for ... _________ min, ...
... but generally fall asleep after no more than ... _________ min.

I prefer to sleep in a completely dark room correct not correct
I wake up more easily when morning light shines into my room correct not correct

How long per day do you spend on average outside (really outside) exposed to day light?
On work days: ___ hrs. ___min. On free days: ___ hrs. ___min.

**Self assessment**

After you have answered the preceding questions, you should have a feeling to which chronotype (time-of-day-type) you belong to. If for example, you like (and manage) to sleep quite a bit longer on free days than on workdays, or if you cannot get out of bed on Monday mornings, even without a Sunday-night-party, then you are more a late type. If, however, you regularly wake up and feel perky once you jump out of bed, and if you would rather go to bed early than to an evening concert then you are an early type. In the following questions, you should categorise yourself and your family members.

Please tick only one possibility!

Description of categories:

extreme early type = 0
moderate early type = 1
slight early type = 2
normal type = 3
slight late type = 4
moderate late type = 5
extreme late type = 6

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **I am**... | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| **as a child, I was** ... | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| **as a teenager, I was** ... | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| **In case you are older than 65:** In the middle of my life, I was ... | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| My parents are/were... | | | | | | | |
| **Mother** ... | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| **Father** ... | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| **My siblings are/were** ... (please underline Brother or Sister) | | | | | | | |
| Brother/Sister | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Brother/Sister | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Brother/Sister | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Brother/Sister | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Brother/Sister | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Brother/Sister | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Brother/Sister | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| **My partner** (girl/boy friend, spouse, significant other) is/was ... | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

**Figure 9:** The complete MCTQ questionnaire. (Roenneberg et al. 2003a)

Figure 1. The complete questionnaire. The original was distributed over two pages including a cover letter informing the volunteers about confidentiality and data handling.

As previously mentioned, there are also personality differences among chronotypes, with larks appearing conscientious, trustworthy and emotionally stable, (Mecacci and Rocchetti,1998). Cognitive performance might also be under the influence of chronotypes, and the morning types coped better with early school start times and performed better (Roberts and Kyllonen, 1999). Lastly, Fabbri and collaborators (2007) reported an involvement of the early circadian type with the “left-thinking”, mainly involved in information processing logically and sequentially and in dealing with verbal, analytic, temporal, and digital materials (Fabbri et al., 2007). A recent study (Hu et al., 2016), focused on self-reported morning individuals, showing a significant association of this chronotype with gender, particularly with female (48.4%, versus 37.7% of morning males). Hu and collaborators’ survey also highlighted the lower probability, for morning persons, to have insomnia, to sleep soundly, to sleep walk and that they could require more than 8 hours of sleep per day (Hu et al., 2016)

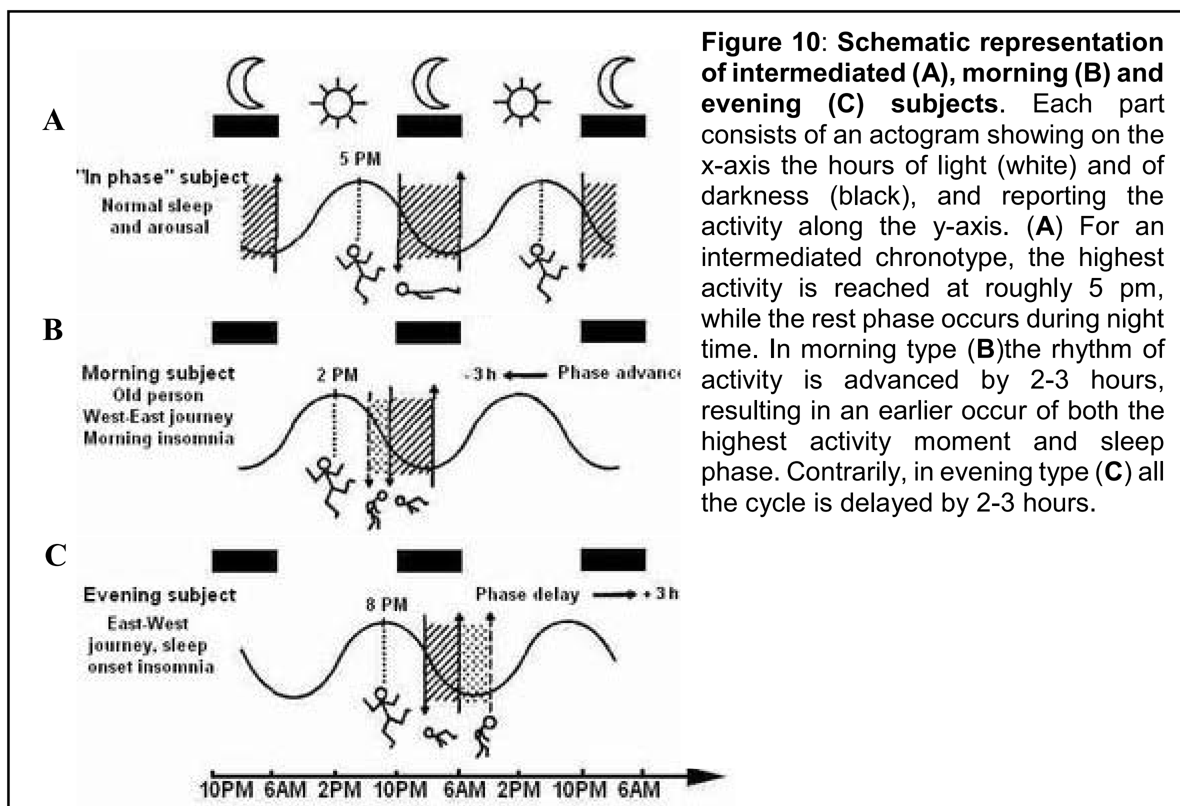



**Figure 10**: **Schematic representation of intermediated (A), morning (B) and evening (C) subjects**. Each part consists of an actogram showing on the x-axis the hours of light (white) and of darkness (black), and reporting the activity along the y-axis. (**A**) For an intermediated chronotype, the highest activity is reached at roughly 5 pm, while the rest phase occurs during night time. In morning type (**B**)the rhythm of activity is advanced by 2-3 hours, resulting in an earlier occur of both the highest activity moment and sleep phase. Contrarily, in evening type (**C**) all the cycle is delayed by 2-3 hours.

On the other hand, the so-called “Evening type” (Figure 10 C), or “owl”, finds it difficult to wake up in the morning and tends to be more active in the second part of the day. It starts the day at a lower body temperature than morning-type subjects, but the temperature increases throughout the day reaching its peak in the late afternoon. Evening chronotypes show a delay in their early-morning peak of salivary cortisol and higher cortisol levels in the first hour after awakening (Kudielka et al., 2006). Those eliciting patterns of behaviour falling in an intermediate area between the two extremes of this continuum are intermediate or

neither-type individuals (Horne and Ostberg 1976). Eveningness has been associated with younger age, being male, having longer sleep duration, but also with being more likely to be affected by depression and using psychiatric medications (Lane et al., 2016). From the personality point of view, while owls were creative, emotionally unstable and had difficult social and familial relations. Despite that, they are reported to be more intelligent (Voinescu, 2009). Finally, evening chronotypes have been scored higher with the right-thinking.Explicitly, the right hemisphere is involved in information processing holistically and nonlinearly and in dealing with nonverbal, concrete, spatial, analogical, emotional, and aesthetic materials (Torrance, 1982 // Fabbri et al., 2007).

### 1.1.5 Polymorphisms associated with particular chronotypes

All the clock, clock-controlled and clock-related genes are important components of the circadian rhythms and mutations at this level could lead to different effect. In this contest, many studies focused on identifying links between clock genes' polymorphisms and their consequences for the organisms, including several diseases, mainly through phenotype-driven mutagenesis screens in animal models, as flies and mice (Buhr and Takahashi, 2013). The first mammal identified as a spontaneous circadian mutant was the tau-mutant hamster which displays a free-running period of 20 h, compared to a wild-type free-running period of 24 hours (Ralph and Menaker 1988). This shortened period results from a mutation in the enzyme casein kinase 1ε (Lowrey et al. 2000). In other two independent studies a chemically induced mutation has been found to be responsible for long-period phenotypes in mice was found in the F-box gene Fbxl3 (Godinho et al., 2007). FBXL3 polyubiquitinates CRY proteins, thereby targeting them for proteosomal degradation (Busino et al., 2007). Interestingly, CRY1 and CRY2 are targeted for ubiquitination by unique phosphorylation events and kinases.

Those experiments, comparing genotypic and phenotypic information, highlighted the involvement of some clock genes' mutations with particular phenotypic alterations. Therefore, there is reason to believe that a great number of serious and less serious disorders may be related with dysfunctions of the circadian clock.

Those evidences represented the starting point for similar investigation in our specie (Katzenberg et al. 1998; Archer et al. 2003; Ko and Takahashi, 2006; Viola et al., 2007; Archer et al., 2008; Langmesse et al. 2008 ; Voinescu et al., 2009; Allebrandt et al., 2010; Pedrazzoli et al., 2010; Utge et al. 2010; Fisher et al. 2013). These studies tried to identify and clarify links between clock gene's polymorphisms and human abnormal phenotypes, include morningness/eveningness preferences. Human preferences in the timing of sleep and wake are, at least partly, based on genetics (Ebisawa et al., 2001; Jones et al., 1999;

Katzenberg et al., 1998; Toh et al., 2001) and, as with other genetic traits, circadian properties depend on specific alleles of clock, clock-controlled and clock-related genes (for review, see Young and Kay, 2001). Differences in circadian rhythms have been associated with medically relevant traits such as sleep, obesity and depression (Zisapel, 2001; Roenneberg et al, 2012; Germain and Kupfer, 2008). However, in my thesis, I considered SNPs that, in light of the present day knowledge, seems to be involved only in circadian rhythms and sleep disorders, excluding polymorphisms shown to be associated with metabolic syndromes and mood disorders.

Although genetic analyses have been done over many worldwide populations, considering both healthy and affected people, the relation between clock genes' polymorphisms and human sleep diseases are still unclear. Indeed, for some polymorphisms there are no concordant evidences for associations genotype at clock genes and circadian phenotypes shared between populations around the word (Barbosa et al., 2010).

The first clock gene polymorphism that was reported to be associated with phenotype in humans was a single-nucleotide polymorphism (SNP) in the 3-untranslated region of *Clock* (T3111C). The minor allele (C) was reported to be associated with an increased evening preference, in a mostly North American population (Katzenberg et al., 1998). This polymorphism was further investigated in other populations. Associations were found also in North-Americans (Friedman et al., 2009) and Japanese (Mishima et al., 2005). Negative results were reported in British (Robilliard et al. 2002), mixed European (Johansson et al., 2003), Korean (Lee et al., 2007) and Brazilian (Pedrazzoli et al., 2007) samples. Carpen and collaborators (Carpen et al., 2006) reported the first evidence of association between a PER1 polymorphism and extreme diurnal preference. A silent polymorphism in exon 18, T2434C, conferred a tendency towards extreme morning preference on carriers of the C allele (frequency = 0.24) than in subjects with extreme evening preference (frequency = 0.12). Investigations on a British sample of PER2 gene highlighted a polymorphism in the 5'-untranslated region (5'-UTR) involved in particular chronotypes. More precisely, the 111G allele frequency was significantly higher in subjects with extreme morning preference (0.14) than in subjects with extreme evening preference (0.03) (Carpen et al., 2005). Similar results have been obtained for another *Per2* SNP (A2114G), which A allele associates with eveningness in a Japanese population (Matsuo et al., 2007).

One of the most interesting sequence variants is a variable-number tandem repeat (VNTR) polymorphism in the human Per3 gene, in relation to diurnal preference or circadian disorders. The SNP encodes 18 amino acids repeating either four times (*Per3* - 4 allele) or five times (*Per3* - 5 allele). In a UK-based population sample, it was found that the shorter allele (*Per3* - 4 allele) was associated with evening preference and delayed phase sleep syndrome, whereas the longer one (*Per3* - 5 allele) with morning preference (Archer et al.,

2003). In contrast to this finding, a study in a Brazilian population found that the frequency of the longer allele was higher in patients with delayed phase sleep syndrome, resulting in a dependence on age and latitude (Pereira et al., 2005; Jones et al., 2007).

### 1.1.6 Abnormal circadian rhythms: FASPS and DSPS

From an evolutionary point of view, the circadian clock represents an advantage for organisms to anticipate and adapt to environmental changes allowing organisms optimize certain behaviours, such as sleep, to occur at designated times. Alteration of the circadian system give rise to the so-called “Circadian rhythm sleep disorders” (CRSDs), defined by persistent or recurrent disturbed sleep–wake cycles (Hida et al., 2014). CRSDs can be categorized according to their postulated underlying mechanisms: 1) the external environment and/or social circumstances are altered relative to the endogenous circadian clock as during jet lag and shift work disorder; 2) the endogenous circadian clock itself is altered, such as irregular sleep wake rhythm, free-running, and disorder sleep-wake state dissociation disorders) (Zhu and Zee, 2012). Among the latest subgroup, there are the extreme morningness which can result in the (familial) advanced sleep phase syndrome (ASPS) and the extreme eveningness resulting in the delayed sleep phase syndrome (DSPS) (Allebrandt and Roenneberg, 2009; Hida et al., 2014).

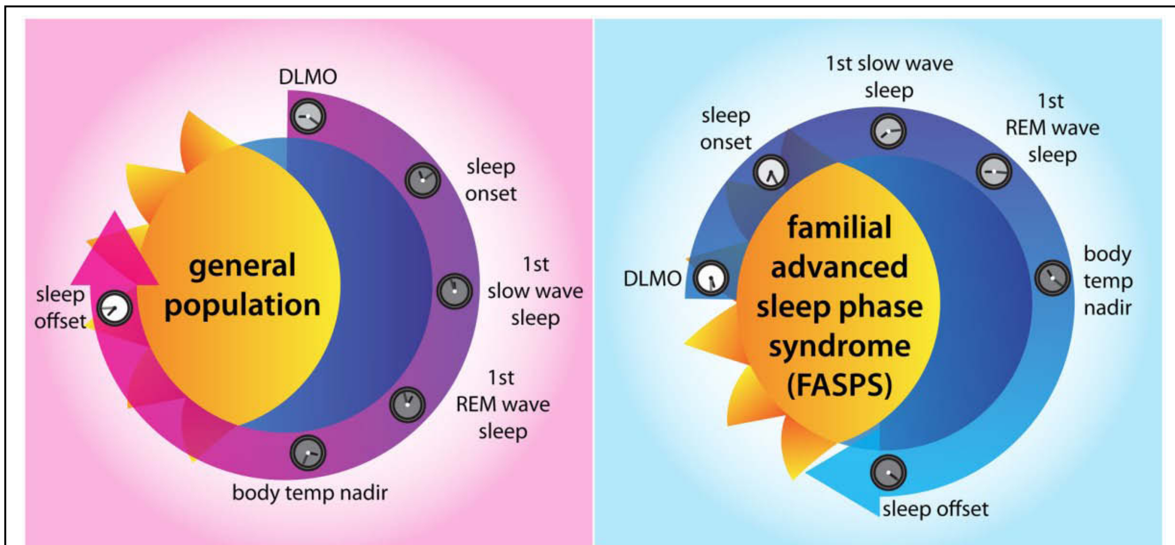
Owing to the complexity of polygenic interactions, confounding environmental factors, and a lack of human relevant experimental models, pioneering work on understanding the regulation of circadian rhythm relied on model organisms, namely zebrafish, fruit flies and worms. With the advent of high-throughput genotyping technologies, and the identification of conserved circadian genes, mounting evidence now supports the role of genetics in regulating sleep timing (Sehgal and Mignot, 2011). The major advances include the identification of molecules regulating sleep and the realization that sleep disorders are extremely common and numerous. Nevertheless, to overcome these difficulties, familial approach represents a useful technique in the field of human genetics in order to find associations between behaviours and genes.

In this contest, families with Mendelian sleep traits have been investigated (Chong et al., 2012). The first recognized human Mendelian circadian rhythm phenotype is the Familial Advanced Sleep Phase Syndrome (FASPS), which presents a highly penetrant autosomal dominant mode of inheritance in three large Utah families (Jones et al., 1999; Reid et al., 2013).

When the Home-Östberg questionnaire, which measures "morning lark" or "night owl" tendency, was performed on clinically affected and unaffected family members, individuals thought to be phase-advanced did score significantly higher than unaffected relatives.

FASPS appeared to be early onset (8 years old), and in affected individuals are characterized by a 3-4 hours phase advance in sleep-wake, melatonin and body temperature cycles (Figure 11; Jones et al., 1999; Reid et al., 2001; Chong et al., 2012).

Therefore, in patients affected by Advanced Sleep Phase Syndrome, sleep onset occurs at approximately 7:30 p.m., when most people are actively socializing. Sleep duration is




**11**: **Clinical phase markers for familial advanced sleep phase syndrome (FASPD)**
FASPD-affected individuals (right) exhibit a marked advancement in phase markers of overt rhythms related to sleep physiology as compared to the general population (left). Melatonin and temperature rhythms are advanced by 4–6 hours, and the free running period has been measured as 1 hour shorter than in controls (Sehgal and Mignot, 2012). Clock faces indicate the mean time of occurrence for each event, with lighter clock faces indicating morning hours and darker clock faces indicating evening hours. Abbreviations: DLMO, dim light melatonin onset. REM, rapid eye movement. Temp, temperature (Chong et al., 2012).

normal, but is terminated by a spontaneous awakening at approximately 4:30 a.m. just when conventional sleepers are at their sleepiest time of the 24-hour cycle (Toh et al., 2001). In order to determine the genetic mechanism underlying FASPS, linkage analysis was performed for one large family, which mapped the allele to chromosome 2qter (Toh et al., 2001).

Underlying mutations pointed to defects in phosphorylation of PER2 as the core issue, with mutations identified in both PER2 and Casein Kinase 1 genes (CK1δ) (Figure 12). Firstly, FASPS has been linked to a mutation in *Per2* gene, at a site normally phosphorylated by *CKIε*. The base change at position 2106 (A-G) of the hPER2 cDNA results in substitution at the amino acid 662 from serine to glycine, PER2-S662G (Jones et al., 1999).

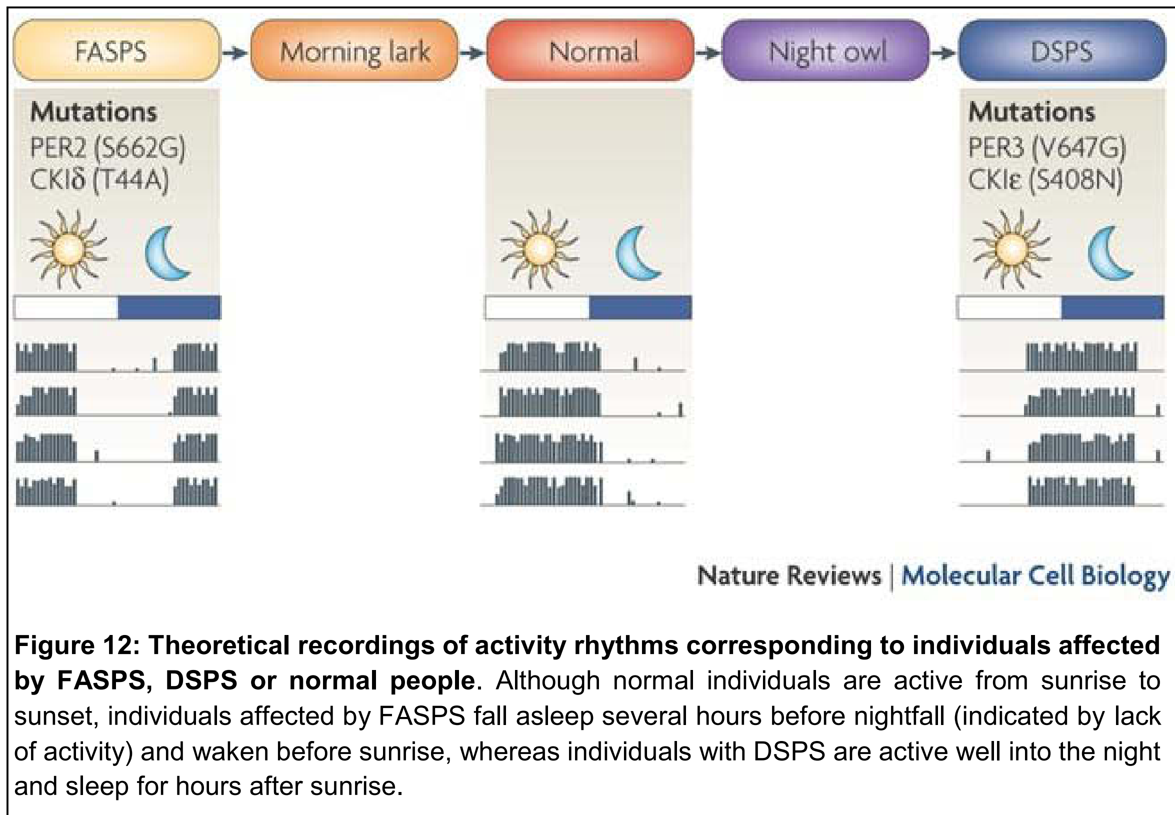



**Figure 12: Theoretical recordings of activity rhythms corresponding to individuals affected by FASPS, DSPS or normal people**. Although normal individuals are active from sunrise to sunset, individuals affected by FASPS fall asleep several hours before nightfall (indicated by lack of activity) and waken before sunrise, whereas individuals with DSPS are active well into the night and sleep for hours after sunrise.

The serine at position 662 in hPER2 is the first of 5 serines spaced three amino acids apart. This 'SxxSxxSxxSxxS' motif is highly conserved in mammalian PER proteins (Toh et al.,2001). Phosphorylation at the 662 site by kinases triggers the phosphorylation of two adjacent serines by CK1δ and CK1ε, which rend the *Per2* protein more stable and reduce the period of the circadian clock. The derived allele (PER2-S662G) has been shown to cause a hypo-phosphorylation of PER2 by the kinase (*CKIε*) and increased translocation from the nucleus to the cytoplasm. These findings correspond well with circadian mutations in other organisms. Experimentally, Toh et co-workers (2001) generated transgenic mice using a human bacterial artificial chromosome (BAC) which carries the cis-acting genomic regulatory elements that can faithfully recapitulate endogenous PER2 expression (Xu et al., 2007). Mice were carrying both wild-type *hPER2* and *hPER2*-S662G allele. Behaviour analysis showed that the S662G transgenic mice exhibited ~2 hours shorter free-running period, while, under 12 hours' light:12 hours dark (12L:12D) conditions, the S662G mice showed ~4 hours phase advance of locomotor activity rhythms, which is almost identical to that of human FASPS subjects that carry this mutation.

The second missense mutation that has been shown causing FASPS occurs at a residue that is conserved from mammalian CKIs through *Drosophila* CKI. In particular, it is an A-to-G change in amino acid 44 in the CK1δ protein (Xu et al., 2005). This leads to a threonine-to-alanine substitution (CK1δ-T44A) which decreases the enzymatic activity of the kinase when tested *in vitro*. In order to characterize the functional consequences of this mutation,

BAC transgenic mice with a human containing either the *hCK1δ* wild-type or the *hCK1δ-T44A* were created (Xu et al., 2005). Under free-running conditions, behavioural period was significantly shorter in the mutant transgenic mice compared to wild-type mice, consistent with the phase-advanced phenotype of the human patients that carry this mutation (Zhang et al., 2011). Interestingly, expression of *hCKIdelta*-T44A in *Drosophila* circadian neurons resulted in longer period compared to expression of wild-type *hCKIdelta* (*hCKI δ* -WT). This suggests that despite the highly conserved nature of individual components of the circadian clock in mammals and flies, the interactions of these components as part of the circadian regulatory network may be different (Xu et al., 2007).

Collectively, the model regarding how CKI acts on PER2 to regulate circadian period starts with the phosphorylation of S662 by a priming kinase. The CKIε phosphorylates the serine residues downstream of S662 on *Per2* gene increasing the level of PER2 protein, while CKIδ likely phosphorylates some other site(s) that result in degradation of PER2. In the case of *Per2*-S662G mutation, the S662 residue can no longer be phosphorylated by the priming kinase, leading to hypo-phosphorylation of the downstream residues by CKI. Therefore, the net effect of CKI on PER2 results in reduced PER2 levels and a shorter period (Zhang et al., 2011). CK1δ levels exacerbate the effects of PER2-S662G on circadian period, as transgenic overexpression of CSNK1D in combination with PER2-S662G increases *Per2* degradation through other putative phosphorylation sites and shortens circadian period even more (Chong et al. 2012). These findings correspond well with circadian mutations in other organisms. For example, in *Drosophila* and the Syrian hamster, mutations in the CK1ε kinase, lead to deficient phosphorylation of PER and changes in circadian period (Sehgal and Mignot, 2011).

Delay Sleep Phase Syndrome (DSPD) is the most commonly diagnosed CRSD, predominantly affecting adolescents, young adults, and insomnia patients (Micic et al., 2015). Indeed, its estimated prevalence is of 0.17% in the general population and 7–16% among adolescents (Zhu and Zee, 2012). Opposite to FASPS, people affected by DSPS show a delay of about 4h in their sleep-wake cycle (Kripke et al., 2008). More precisely, individuals sleep (03:00 – 06:00 h) and wake (10:00 – 14:00 h) later than desired (Lowrey et al., 2004). As a consequence, they may be unable to fall asleep early enough to rest adequately before it is time for school or work, or they may be quite unable to report to school or work on time. Consequently, DSPD can be a disabling and socially isolating condition, unless the patients are able to fit their habits into an accommodating social milieu (Kripke et al., 2008). At the molecular level, the pathogenesis of DSPS has been linked, at least, to two SNPs (Figure 12).

The first polymorphism is in the human PER3 gene and induces a Valine-to-Glycine change in amino acid 647 (V647G) (Ebisawa et al, 2001). This residue locates in a region

similar to the CKIε-binding region of PER1 and PER2, close to the serine residue in PER2 that is disrupted by the FASPS mutation. Therefore, this polymorphism might alter the CKIε-dependent phosphorylation of human PER3. A change in the phosphorylation of *Per3* may change its function thus altering the complex cellular machinery (Dauvilliers et al. 2005).

The second mutation which is present in the normal population but absent in individuals with DSPS is in human CKIε. This variant occurs in amino acid 408, inducing a Serine-to-Asparagine substitution (S408N), which is significantly less common in DSPS patients than in control individuals. The protective function is postulated to arise from an alteration in CKIε activity, as it eliminates a putative autophosphorylation site in the autoinhibitory domain of the kinase and leads to a more active kinase, at least in vitro (Takano et al., 2004). Finally, Archer and co-workers (Archer et al. 2003) highlighted the strong correlation between the shorter allele of the length polymorphisms in *Per3* (*Per3*-4) and the DSPS subjects, 75% of whom were homozygous (Lowrey et al. 2004).

## *1.2 Elements of population genetics*

### 1.2.1 The origins of Population Genetics

Population genetics is a field of biology that studies the genetic composition of biological populations, namely the amount and distribution of genetic variation in populations and species. Furthermore, it investigates the various factors that give rise to changes in the genetic composition of populations, including natural selection.

Population geneticists pursue their goals by developing abstract mathematical models of gene frequency dynamics, trying to extract conclusions from those models about the likely patterns of genetic variation in actual populations, and testing the conclusions against empirical data. This field came into being in the last century, thanks to the integration of the previously incompatible of Darwinian Theory with Mendel's idea of particulate inheritance (Jobling et al., 2004). Therefore, population genetics is intimately bound up with the study of evolution and natural selection.

This discipline could be seen as the main achievement of the works of three scientists Fisher R.A., Haldane J.B.S. and Wright S., performed between 1920s and 1930s (Okasha, 2012). Indeed, each of them developed formal models in order to investigate how natural selection, and other evolutionary forces (i.e. mutation), would modify the genetic composition of Mendelian populations over time (Okasha, 2012). Among their studies, the first significant milestone was Fisher's paper (Fisher, 1918) in which he showed how the biometrical and Mendelian research traditions could be unified. Fisher demonstrated that if a given continuous trait, e.g. height, was affected by a large number of Mendelian factors,

each of which made a small difference to the trait, then the trait would show an approximately normal distribution in a population. Since the Darwinian process was widely believed to work best on continuously varying traits, showing that the distribution of such traits was compatible with Mendelism was an important step towards reconciling Darwin with Mendel.

Patterns of genetic variation within and between species reveal the mechanisms and history of evolution. The basic processes of evolution are changes in gene frequency and the emergence of new types by evolutionary forces (including mutation and gene duplication). More precisely, evolution studies the mechanisms of heredity and of variation in allele frequencies, and it concerns with changes in population rather than in individuals.

Before molecular genetics became established, evolutionary biologists had to make inferences based on phenotypic observations. Molecular genetics has provided the means of assessing the genetical biochemistry behind outward phenotypic differences. Nowadays researchers in this area combine molecular, computational, and mathematical approaches to understand how mutation, recombination, gene flow, and natural selection shape the evolution of genomes and ecologically important traits.

Compressively, this discipline describes the behavior of alleles in populations by focusing on the forces that can cause allele frequencies to change over time. "Allele frequency change over time" is simply a definition of "evolution". Therefore, population genetics is that branch of genetics that is concerned with the evolutionary processes of natural selection, genetic drift, mutation, migration, and non-random mating. Population genetic approaches can be used to understand the consequences of these processes individually or in combination.

### 1.2.2 The genetic variability and HWE

The collection of all the alleles of all of the genes found within a freely interbreeding population is known as the gene pool of the population. Each member of the population receives its alleles from other members of the gene pool (its parents) and passes them on to other members of the gene pool (its offspring).

The beginning of the analysis on both allele, so genotypes, frequencies and of the effects of the Mendelian scheme on populations of interbreeding individuals date back to the very beginning of $20^{th}$century. In 1902 Yule (1902) pointed out that, if the members of an F2 population, segregating for a single pair of genes (*A* and *a*), interbreed at random, the three genotypes (*AA*, *Aa* and *aa*) will be present in the same proportions in the following generations. Further independent studies made by G.H. Hardy (an English mathematician; Hardy, 1908) and W. Weinberg (a German physician; Weinberg, 1908) led to postulate, in

1908, the simplest and most important principles in population genetics. The Hardy-Weinberg principle is about the relationship between the allelic and genotypic frequencies and, more precisely, it predicts the expected genotype frequencies using the allele frequencies in a diploid Mendelian population.

Given diploid organisms, the two alleles, *A* and *a*, at the same locus, respectively with frequency $p$ and $q$, can be combined in order to make three genotypes. These will be present in the succeeding generation, accordingly with the following proportions: $AA = p^2$, $Aa = 2pq$, and $aa = q^2$. When the genotype proportions in the succeeding generation is maintained, no evolution (defined as change in allele frequencies) is occurring; therefore, the population is at Hardy-Weinberg equilibrium. However, in order to preserve the proportion through time, the firsts strict valid conditions to be fulfilled are that population size is effectively infinite and characterized by random mating. That is, the absence of a genotypic correlation between mating partners, i.e. the probability that a given organism mates with an *AA* partner, for example, does not depend on the organism's own genotype, and similarly for the probability of mating with a partner of one of the other two types (Okasha, 2012). Furthermore, that random mating will lead the genotypes to be in the above proportions (so-called Hardy-Weinberg proportions) is a consequence of Mendel's law of segregation. To see this, note that random mating is in effect equivalent to offspring being formed by randomly picking pairs of gametes from a large gamete pool, which contains all the successful gametes of the parent organisms. By the law of segregation, a heterozygote (*Aa*) produces gametes bearing the *A* and *a* alleles in equal proportion. Therefore, the relative frequencies of the A and B alleles in the gamete pool will be the same as in the parental population, namely p and q respectively. Given that the gamete pool is very large, when we pick pairs of gametes from the pool at random, we will get the ordered genotypic pairs {*AA*}, {*Aa*}, {*aA*}, {*aa*} in the proportions $p^2:pq:qp:q^2$. Disregarding the order, we can consider the {*Aa*} and {*aA*} pairs as equivalent, giving the Hardy-Weinberg proportions for the unordered offspring genotypes (Okasha, 2012). Importantly, whatever the initial genotypic proportions are, random mating will automatically produce the offspring in Hardy-Weinberg equilibrium in just one generation.

However, for the idealized population to be at HWE, other conditions are required, including the absence of four factors that might change allele frequencies: migration into or out of the population, random genetic drift, mutation and selection. These are rather stringent requirements, and it may be doubted if they are ever fully met; nevertheless, they are often approximated closely enough to make the formula useful in analysis of population.

### 1.2.3 Genetic diversity, admixture and population structure

Human genus originated and evolved in Africa for millions of years. Most evidences support the idea of a recent (<150 KYA) origin, in the same region, of modern humans. The dispersal of groups of individuals "out of Africa" and, more recently, events of migrations (for example due to invasions) brought back into contact previously isolated ancestral populations, allowing the exchange of individuals. The common genetic consequence of the meeting of populations is known as admixture and has led to the generation of hybrid populations (Jobling et al., 2004).

From a genotypic point of view, when individuals from different groups admix, their offspring's DNA becomes a mixture of the DNA from each ancestral group. Pieces of this DNA are passed along through subsequent generations. Consequently, the genomes of admixed individuals contain segments of DNA inherited from each of the original source groups (Figure 13). Therefore, when we analyse the genomes of modern day individuals (who descend from this admixed population) we detect not only the proportion of admixture established during populations first meeting, but the summation of cumulative gene flow from when they first met to the present day (Jobling, et al., 2004).

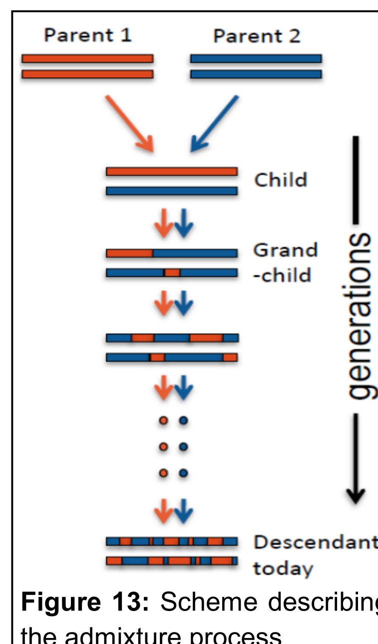



**Figure 13:** Scheme describing the admixture process

Admixed populations are characterized by allele frequencies that are linear combinations of the allele frequencies in their parental populations, and since admixture affects all loci equally, the same set of admixture proportions are expected to apply to all alleles at all loci (Long, 1991).

Of course, other evolutionary forces, including genetic drift, selection and mutations, have also modified the imprint of past admixture in modern populations. In unison, these processes shape all the genetic diversity. Biologically, the impact of admixture is reflected by a variety of phenotypic effects. In particular, quantitative traits that are genetically encoded and well differentiated between populations will be altered in admixed populations. Among them, some example are skin colour, hair colour and stature.

When populations undergo inbreeding, selection or important migration, there is a deviation form Hardy-Weinberg proportions, or deviations from panmixia. In these conditions, populations can be said to be structured in some way, giving rise to a, more or less articulated, population structure. More precisely, it means that, instead of a single, simple population, populations are somehow subdivided. The overall "population of

populations" is often called a meta-population, while the individual component populations are often called subpopulations.

Several studies (Barbujani et al. 1997; Jorde et al. 2000) postulated that roughly 85 – 90 % of neutral genetic variation in the human species is due to differences between individuals within populations. The remaining 10%–15% is distributed between groups, and, though modest, this variation influences the average differences in physical characteristics, disease susceptibility, and treatment outcome among populations (Bamshad et al., 2003).

To assess the impact of this variation, particularly in comparison with environmental factors, inferences are often made about the genetic structure of a sample (e.g., the number of subpopulations) and about which individuals are assigned to each subpopulation. This is important because it may be better to consider each subpopulation separately in some situations (e.g., testing whether there is any difference between groups in the effects of natural selection or genetic drift). Thus, a major goal of population genetics is to understand the nature and extent of human population structure.

Inference on genetic ancestry differences among individuals from different populations, that is, the population structure, has been motivated by a variety of applications, ranging from genetic association studies to personalized medicine, forensics and evolutionary study. In the latter contest, indeed, researchers try to learn about the relationship between the ancient people and modern populations in order to infer the evolutionary history of human beings (Liu and Zhao, 2006). In order to identify past admixture excluding selection process as explanation for modern patterns genetic diversity, a focus on noncoding or neutral genetic markers is required. Therefore, modern genetic data combined with appropriate statistical methods have the potential to contribute substantially to understanding our past (Hellenthal et al, 2014).

This has been possible also thank to the great advancements in array-based genotyping technologies, which have largely facilitated the investigation of genetic diversity at remarkably high levels of detail. Indeed, in the last decades, technologies such as high-density genotyping arrays and next generation resequencing have facilitated the production of an enormous quantity of data with which to investigate genetic relationships in humans and in other organisms (Lawson et al., 2012). Consequently, a variety of methods has been proposed for the identification of genetic ancestry differences among individuals in a sample using high-density genome-screen data. For this purpose, in my thesis I applied two of the most popular approaches to investigate population structure through genetic data: the Principal Components Analysis and the estimation of admixture memberships. It is common to apply both approaches to the same dataset, in order to provide a useful summary of the basic features of the data (Lawson et al., 2012).

### 1.2.4 Mutation generates genetic diversity

Mutation is the ultimate source of genetic diversity, and it is defined as the change in DNA sequence. It includes a large broad of events, with different rates and different molecular mechanisms. Among them, there are insertions and deletions of few bases, duplications and inversions of mega base segments of DNA, and translocation of chromosomal segments. The simplest difference between two homologous DNA sequences is the base substitution, in which one base is exchanged with another. It gives rise to the so-called single nucleotide polymorphisms (SNP). These represent the molecular marker investigated in my thesis.

### *1.2.5 Natural selection*

As humans and other organisms moved to inhabit every part of the world, they were exposed to many new and different environments, diets, and pathogens. All these external conditions, forced different species to adapt, leading to the great diversity we observe today. In population genetics, the ability of an individual to survive and reproduce relative to the rest of the population is referred as fitness, and is partly dependent on the environment.

In 1858, Darwin and Wallace laid the foundation for the species evolution, postulating the principle of natural selection, based on the observation that fitness-enhancing traits tend to become more frequent in populations over time (Vitti et al., 2013). Accordingly, natural selection represents the differential reproduction of genotypes in succeeding generations. It can occur at any stage from fertilization to generation of progeny and takes place when some genotypic variant in a population enjoy a survival or reproduction advantage over other (Okasha, 2012). Consequently, the selective pressure produces individuals with different fitness in different environments (Vitti et al., 2013).

Searching for evidence of adaptation and for specific changes that underling it, Haldane (1949) uncovered the first adaptive trait in human. He observed that many diseases of red blood cells seemed to be distributed in regions where malaria was endemic. Inferring adaptation is no trivial task (Li et al. 2012). However, through evolutionary genetics analysis, many adaptive traits have been elucidated. Among them, the armored plates in stickleback fish (Jones et al., 2012) and the coat colour in field mice (Mullen et al., 2009).

Only for a handful of human loci there are convincing evidences, from both genomic analyses and in vitro or in vivo studies of functional differences between genotypes, that some well-identified environmental factors shaped patterns of variation. Examples include skin pigmentation (Parra, 2007), the ability to digest lactose (Bersaglieri et al. 2004), and the malaria-related polymorphisms at the Duffy blood group (Hamblin and Di Rienzo 2000)

and G6PD (Tishkoff et al. 2001) loci. For many other loci, evidence suggestive of positive selection pressures does exist, but the patterns in the data are also consistent with neutral evolutionary mechanisms related with demographic history (Harris and Meyer 2006), i.e. the processes leading to the anatomically modern humans' dispersal over the whole planet and successive local events.

In the genomic era, selection refers to any non-random, differential propagation of an allele as a consequence of its phenotypic effect. In the past decades, many models have been proposed in order to describe how selection could occur. Generally, mutation which increase fitness are referred as undergo positive selection. The involved alleles are favoured and, therefore, propagated. On the other hand, mutations that reduce the fitness of the carrier are subject to negative (or purifying) selection, and acts on disfavoured allele (Vitti et al., 2013). In either cases, selection causes a form of "bottleneck" that is limited to the selected site and is characterized by a local loss of genetic variation.

Random mutations are more likely to be deleterious than beneficial, so many novel alleles are immediately subject to negative selection and become removed from the gene pool before they can achieve detectable frequency within the population. This ongoing removal of deleterious mutations is a form of negative selection referred to as background selection. In genetic regions under strong background selection, mutations are quickly removed from the gene pool, resulting in highly conserved stretches of the genome.

More subtle configurations of positive and negative selection give rise to other common evolutionary trends, particularly (although not exclusively) in diploid and polyploid organisms, where the phenotype depends on the interaction of multiple alleles at the same locus. Among them, there is balancing selection, in which multiple alleles are maintained at an appreciable frequency within the gene pool. If the alleles being maintained conduce to opposing phenotypic effects — for example, if large and small body sizes are maintained within the population to the exclusion of intermediate sizes — then the trend is often further described as diversifying or disruptive selection. By contrast, when intermediate phenotypic values are favoured, whether by balancing selection of codominant alleles or by positive selection of alleles that underlie intermediate phenotypes, the trend is called stabilizing selection (Vitti et al., 2013).

Importantly, positive selection is considered be the main mechanism of adaptation, namely the genesis of phenotypes that are proper for a specific environment. For this reason, in my thesis, I focused the attention on the investigation of positive selective pressure within our species. Indeed, positive selected variants are thought to be indicative of local adaptation that characterized human evolution and history.

Local adaptation to divergent environments can lead to deep differences in average phenotype between populations of the same species. Such variation offers particularly convincing evidence of natural selection when it is correlated with variation in ecological factors over multiple independent geographic regions and/or species. Correlations between phenotype and natural environment may be reflected at the individual level, where at some loci, allele frequencies strongly differentiate populations that live in different ecological niche. Such correlations can arise when selection pressures exerted by the environmental variable are sufficiently divergent between geographic locations, such that differences in allele frequency can be maintained in the face of gene flow.

The advent of genome-wide data sets, characterized by individuals from many populations across a wide geographic range, (e.g. Nordborg et al., 2005; Jakobsson et al., 2008; Li et al., 2008; Auton et al., 2009) allows researchers both to obtain a systematic view of the processes shaping local adaptation and to gain valuable insights into the genetic and ecological basis of adaptation and speciation (Coop et al., 2010).

In order to investigate evidences of natural selection acting in our species, I applied different approaches based on summary statistics which compare empirical data with expectations under the null hypothesis of selective neutrality. Neutrality tests are, indeed, based on Kimura's neutral theory (1985), which postulates that the vast majority of genetic change is attributable to genetic drift, namely a random fluctuation of allele frequencies in a population due to chance in the contribution of each individual to the next generation, rather than Darwinian selection.

## 2. AIMOF THE THESIS

As previously described, the main function of biological clocks is to provide temporal information to the organism, in order to coordinate physiological and/or behavioral responses during the daily cycle so that they can maximize their adaptation. Environmental parameters vary dramatically as a function of latitude and locale, creating specific selective pressure (Hut et al., 2013). Therefore, biological clocks within wide-ranging species may be genetically "tuned" by natural selection to optimize adaptiveness along latitudinal clines or other location dependent factors. In wide-ranging species other than humans, natural selection has genetically optimized adaptiveness along latitudinal clines (Ciarleglio et al., 2008). A well-studied example is represented by *Drosophila melanogaster*, which exhibits genetic polymorphisms in circadian clock genes along latitudinal clines (Sawyer et al. 2006; Tauber et al. 2007). *Homo sapiens* is a species that is distributed to practically every terrestrial niche on Earth. Therefore, it is likely that peculiar geographic diversity patterns in circadian genes might represent an adaptive response to different light/dark cycles or environmental changes to which different human populations are exposed (Ciarleglio et al., 2008). However, results from previous studies of classical markers in clock genes (for example, CLOCK-T3111C and PER3-VNTR) are not completely consistent for different populations around the world (Archer et al., 2003; Barclay et al., 2011; Chang et al. 2011; Katzenberg et al., 1998; Lazar et al. 2012; Osland et al. 2011; Pereira et al., 2005; Robilliard et al., 2002). Investigations on novel candidate SNPs in clock genes, in order to achieve a deeper understanding of the genetic basis of circadian phenotypes in humans, are fundamentals.

In this contest, the main purposes of my thesis were (1) to assess the nature of the evolutionary mechanisms that could have played a role in shaping the patterns of variation in circadian clock genes observed in present-day human populations, and (2) to identify environmental parameters of relevance to biological clocks.

To do so, we firstly assembled a large database of DNA polymorphisms located in human clock, clock-controlled and clock-related genes, and used populations genetic analysis in order to compare observed data with expectations generated under the null hypothesis of selective neutrality. In such a way, we could identify loci in which populations sharply differ from each other. More precisely, through a candidate gene approach, we investigated SNPs in order to highlight loci showing evidences of positive selection, namely local adaptations and/or clinal adaptation.

Secondly, to investigate the hypothesis that genetic variation in components of the human circadian clock has been shaped by environmental variables, such as photoperiod and/or temperature profiles, we performed correlation analyses testing the relationship between

allele frequencies of SNPs known to be associated with sleep disturbances and diseases (at least in one human population) and some important ecological factors.

## 3. MATERIALS AND METHODS

### *3.1 Genomic data: the 1000Genome Project*

Recent improvements in sequencing technology, as an example the "next gen" sequencing platform, have sharply reduced the cost of sequencing and rapidly increased the amount of genetic data available.

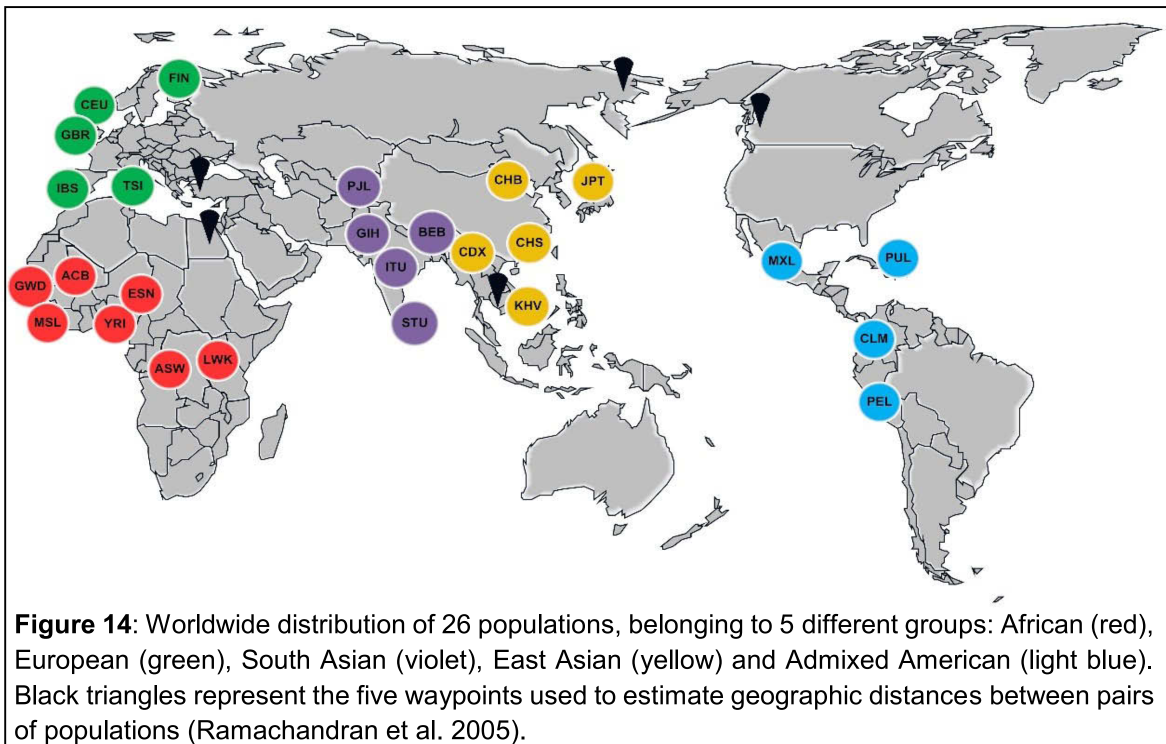



**Figure 14**: Worldwide distribution of 26 populations, belonging to 5 different groups: African (red), European (green), South Asian (violet), East Asian (yellow) and Admixed American (light blue). Black triangles represent the five waypoints used to estimate geographic distances between pairs of populations (Ramachandran et al. 2005).

The 1000 Genomes Project, began in 2007, has been the first project to sequence the genomes of a large number of people, with the goal of developing a comprehensive description of common human genetic variation by applying whole-genome sequencing to a diverse set of individuals from multiple populations (1000 Genomes Project Consortium et al. 2015). The final phase 1000 Genomes Project publications represent not only the completion of this project, but also the culmination of a series of international collaborations stemming from the HGP, including the International HapMap Project, all focused on establishing open reference catalogues of common and rare variation, providing insights into the processes that shape genetic diversity, and advanced understanding of disease biology as a resource to the community.

From the 1000Genome browser (http://www.1000genomes.org/; 1000 Genomes Project Consortium et al. 2015), I collected genetic data for SNPs genotyped in 2,504 individuals, belonging to 26 worldwide populations. Among the populations considered, 7 are from Africa, 5 from Europe, 5 from South Asia, 3 from East Asia and 4 are defined as Admixed American (Figure 14, Table 1).

| Population | Population Description | Super Population | Size | Latitude | Longitude |
|---|---|---|---|---|---|
| a_ACB | African Caribbeans in Barbados | AFR | 96 | 13,12 | -59,62 |
| a_ASW | Americans of African Ancestry in SW USA | AFR | 61 | 0,00 | 24,00 |
| a_ESN | Esan in Nigera (*Edo State*) | AFR | 99 | 6,50 | 6,00 |
| a_GWD | Gambian in Western Divisions in The Gambia | AFR | 113 | 13,27 | -16,65 |
| a_LWK | Luhya in Webuye, Kenya | AFR | 99 | 0,62 | 34,77 |
| a_MSL | Mende in Sierra Leone (*Bo district*) | AFR | 85 | 8,00 | -11,67 |
| a_YRI | Yoruba in Ibadan, Nigera | AFR | 108 | 7,40 | 3,92 |
| b_CEU | Utah Residents (CEPH) with Northern and Western European ancestry | EUR | 99 | 48,86 | 2,35 |
| b_FIN | Finnish in Finland | EUR | 99 | 60,17 | 24,933333 |
| b_GBR | British in England and Scotland | EUR | 91 | 52,83 | -3,86 |
| b_IBS | Iberian population in Spain | EUR | 107 | 40,43 | -3,68 |
| b_TSI | Toscani in Italia | EUR | 107 | 43,77 | 11,25 |
| c_BEB | Bengali in Bangladesh | SAS | 86 | 24,00 | 88,00 |
| c_GIH | Gujarati Indian from Houston, Texas (*Gujarat*) | SAS | 103 | 23,22 | 72,68 |
| c_ITU | Indian Telugu from the UK | SAS | 102 | 16,94 | 79,56 |
| c_PJL | Punjabi from Lahore, Pakistan | SAS | 96 | 31,55 | 74,34 |
| c_STU | Sri Lankan Tamil from the UK (*Tamil Nadu*) | SAS | 102 | 13,09 | 80,27 |
| d_CDX | Chinese Dai in Xishuangbanna, China | EAS | 93 | 22,00 | 100,80 |
| d_CHB | Han Chinese in Bejing, China | EAS | 103 | 39,91 | 116,39 |
| d_CHS | Southern Han Chinese | EAS | 105 | 27,07 | 115,66 |
| d_JPT | Japanese in Tokyo, Japan | EAS | 104 | 35,69 | 139,69 |
| d_KHV | Kinh in Ho Chi Minh City, Vietnam | EAS | 99 | 10,77 | 106,68 |
| e_CLM | Colombians from Medellin, Colombia | AMR | 94 | 6,29 | -75,536111 |
| e_MXL | Mexican Ancestry from Los Angeles USA | AMR | 64 | 19,43 | -99,13 |
| e_PEL | Peruvians from Lima, Peru | AMR | 85 | -12,04 | -77,03 |
| e_PUR | Puerto Ricans from Puerto Rico | AMR | 104 | 18,45 | -66,07 |
| 26 Populations | | 5 Superpopulations | **2504** | | |

**Table 1: List of 26 populations included in the study.** For each population, are reported the sample size, latitude and longitude. AFR, African; EUR, European; SAS, South Asian; EAS, East Asian; AMR, Admixed American

I assembled three genetic datasets: two consisting of clock genes' polymorphisms and one with presumably neutral SNPs. The first includes all SNPs in 25 autosomal genes, namely clock, clock-controlled and clock-related genes (Table 2).

| Gene | Chromosome | Position |
|---|---|---|
| ***Positive elements of the Molecular Clock*** | | |
| BMAL1 | 11 | 13248199 - 13458813 |
| CLK | 4 | 56244070 - 56463305 |
| NPAS2 | 2 | 101386614 - 101663291 |
| RORα | 15 | 60730483 - 61571518 |
| RORß | 9 | 77062281 - 77352117 |
| RORγ | 1 | 151728547 - 151854348 |
| ***Negative elements of the Molecular Clock*** | | |
| DEC1 | 9 | 117854097 - 118214923 |
| DEC2 | 12 | 26222959 - 26328060 |
| CRY1 | 12 | 107335142 - 107537607 |
| CRY2 | 11 | 45818669 - 45954798 |
| PER1 | 17 | 7993790 - 8109824 |
| PER2 | 2 | 239102679 - 239248743 |
| PER3 | 1 | 7794380 - 7955237 |
| REV - ERBα [NR1D1] | 17 | 38199040 - 38306978 |
| TIM | 12 | 56760903 - 56893187 |
| ***Clock-Controlled and Clock-related Genes*** | | |
| AANAT | 17 | 74399433 - 74516199 |
| AVPR1A | 12 | 63489014 - 63594722 |
| AVPR1B | 1 | 206173976 - 206281639 |
| CKIδ | 17 | 80146899 - 80281607 |
| CKIε | 22 | 38636697 - 38844527 |
| CKII | 20 | 409116 - 574482 |
| DBP | 19 | 49083287 - 49190695 |
| MTNR1A | 4 | 187404809 - 187526721 |
| MTNR1B | 11 | 92652886 - 92768232 |
| OPN4 | 10 | 88364314 - 88476605 |

**Table 2:** List of 25 candidate autosomal genes included in the analyses. For each gene is reported the chromosome and the position (in bp), Clock genes are classified as "positive" and "negative" elements depending on their role in the loops. Clock-controlled and clock-related genes are elements associated to the circadian clock mechanism

In order to include the 5' and 3' flanking regions, I included also 50 bp at the beginning and at the end of the gene. Out of the set of 135,162 SNPs, I retained a subset of polymorphisms showing a level of pairwise Linkage Disequilibrium lower than r2=0.3 (as estimated in the entire dataset) by the PLINK v1.07 tool (Purcell et al. 2007). I applied to the whole dataset the following command line: "*plink --file data --indep-pairwise 50 5 0.3",* which prunes polymorphisms within a sliding window based on pairwise genotypic correlation. We chose a sliding window size of 50 SNPs, a shift step of 5 SNPs and a r2 threshold of 0.3. This pruned subset contains 116,440 SNPs and it is referred as the Candidate dataset.

Among these SNPs, I identified in the literature 15 polymorphisms belonging to 11 clock genes (Table 3), for which was reported an association with human sleep disorders, such as the Advanced Sleep Phase Syndrome (ASPS) and the Delayed Sleep Phase Syndrome (DSPS), or with changes in circadian phenotypes, at least in one world population. The dataset including phenotypic information, referred to as "Phenotype-associated" dataset, is not independent form the Candidate dataset, but no analysis were run in parallel on both.

| Gene | SNP | Reference |
|---|---|---|
| | ***Positive Clock Genes*** | |
| CLK | rs11932595 | Allebrandt et al. 2010; |
| | rs6850524 | Allebrandt et al. 2010; |
| NPAS2 | rs7598826 | Allebrandt et al. 2010; |
| | ***Negative Clock Genes*** | |
| PER1 | rs2735611 | Carpen et al. 2005; von Shantz et al. 2008; |
| PER2 | rs2304670 | von Shantz et al. 2008; |
| | rs934945 | Lee et al. 2011; |
| PER3 | rs35426314 | Ciarleglio et al. 2008; von Shantz et al. 2008; |
| | rs2797687 | Archer et al. 2010; |
| | rs228730 | Archer et al. 2010; |
| | rs2640909 | Ojeda et al. 2013; |
| TIM | rs2291738 | Allebrandt et al. 2010; |
| | ***Clock-Controlled and Clock-related Genes*** | |
| AANAT | rs28936679 | Hohjoh et al. 2002; Barbosa et al. 2010; |
| CKIδ | rs7209167 | Allebrandt et al. 2010; |
| CKIε | rs77945315 | Takano et al. 2004; Barbosa et al. 2010; |
| OPN4 | rs2675703 | Roeckelin et al. 2012 |
| REVERB | rs12941497 | Kang et al. 2015 |

**Table 3:** List of SNPs included in the "Phenotype-associate" dataset. The 15 polymorphisms, with known link in the literature to human sleep disturbances or changes in circadian phenotypes and in linkage equilibrium, belong to 11 clock genes. For each locus we report the reference to previous studies.

Finally, I created a Reference dataset using the NRE software (http://nre.cb.bscb.cornell.edu/nre/, Arbiza et al. 2012) applying Patin et al. (2009) criteria. It consists of 200,000 autosomal neutral sites, sampled at random in non-coding regions and at least 200 kb away from any known or predicted gene.

### *3.2 Environmental data*

There is a close relationship between genotype and phenotype. Most complex traits, such as hair or skin colour, height, weight or behaviour, are influenced by many genes. Most traits are also influenced by environment. This means that the same genotype can result in different

phenotypes, depending on the environment. For example, between two people with a genetic risk for lung cancer, for the one that smokes is much more likely to develop the disease. Environmental effects also imply that the same phenotype can result from more than one genotype; smoking again provides an example, because most smokers who are not genetically at risk can also develop lung cancer (Voinescu, 2009).

Among the environmental parameters of relevance for biological clocks, the main factors seems to be (i) daily fluctuations in both light intensity and temperature, and (ii) seasonal changes in photoperiod (daylength) and temperature; these parameters vary dramatically as a function of latitude and locale (Ciarleglio et al., 2008).

In my thesis, I included in the analyses 7 environmental factors, i.e. photoperiod, humidity, radiation fluxes, precipitation, temperature and two proxy variables summarizing climatic diversity: latitude and longitude. I also considered the geographic distance from an arbitrary point in East Africa (Addis Ababa) roughly representing the origin of the expansion processes of anatomically modern humans, thus taking into account one crucial demographic event affecting the allele frequency distribution.

Geographic distances form Addis Ababa were calculated as great circle distances. In agreement with previous studies (Ramachandran et al., 2005), to better reflect in those distances the likely routes of human dispersal between continents, distances between populations of different continents were calculated through five obligatory waypoints (across the Suez Canal, between Anatolia and Europe, in Northeastern Russia, Cambodia and Northwestern Canada).

To test for any effects of the seasonal variation of the photoperiod, we calculated for each population the difference in photoperiod between the Summer and the Winter solstices (http://aa.usno.navy.mil/data/docs/RS_OneYear.php.).

The NCEP/NCAR Reanalysis Project at the NOAA/ESRL Physical Sciences Division (http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml) contains raw data of humidity and solar radiation fluxes starting from January 1$^{st}$ 1948 until January 1$^{st}$ 2015. For each population, the mean relative humidity was calculated as the ratio between the actual vapour density and the saturation vapour density. Regarding the solar radiation fluxes, we took into consideration the net shortwave (nswrs) and longwave (nlwrs) radiation fluxes. The former represents a measure of the difference between the incoming solar shortwave radiation and the outgoing shortwave radiation from the earth surface, whereas the latter is a measure of the difference between the outgoing longwave radiation from the earth surface and the incident atmospheric longwave counter-radiation. For each sample, in reference to the period January 1$^{st}$ 1948 – January 1$^{st}$ 2015, we calculated the ratio between the average of nlwrs and of nswrs.

Information for both the temperature and the precipitation rate have been obtained from the

Climatic Research Unit (CRU) database (http://www.cru.uea.ac.uk/). Monthly average temperatures were measured starting from 1850, in stations at different elevations, often using different methods. In order to avoid biases that could result from these problems, temperatures in the database were reduced to anomalies from the period with best coverage, 1961 – 1990. I calculated, for each population, the mean annual temperature, considering the number of years for which data are available. Conversely, data about precipitations 1900 – 1998 are available for all 26 populations investigated in this study. For each of them, I estimated the mean precipitation rate, expressed as mm*10.

Further, we included in our analysis the longitude and latitude of each population. The latter is a common proxy for other, unspecified environmental variables, which may have an effect of climate and/or photoperiod (Gunther and Coop, 2013).

## *3.3 Analysis of population structure*

### 3.3.1 Principal Component Analysis

Multivariate data consist of observation on several different variables for a number of individuals or object. These data type is arising in all branches of science, ranging from psychology to biology. As a consequence, at the present day, a huge number of multivariate techniques are available (Chatfield and Collins, 1980). The underlying theme of much of these analyses is simplification. In other words, multivariate data analysis refers to any statistical technique used to analyse data that arises from more than one variable and allow researchers to reduce the multidimensional space, while reducing the inevitable loss of information. Principal Component Analysis (PCA) is a commonly used example of these approaches.

PCA is a one-sample technique applied to data with no groupings among the observations and no partitioning of the variables into subvectors $y$ and $x$. Principal components are concerned only with the core structure of a single sample of observations on p variables (Rencher, 2002). Its goal is to extract the important information from the data, to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations and of the variables as points in maps, namely a scatter plot (Abdi and Williams, 2010). Principal components are orthogonal because they are formed with eigenvectors of the covariance (or correlation) matrix, which is symmetric. Generally, in principal component analysis, we seek to maximize the variance of a linear combination of the variables. The first principal component is the linear combination with maximal variance. The second principal component is the linear combination with maximal variance in a direction orthogonal to the first principal component, and so on. The first principal component

also represents the line that minimizes the total sum of squared perpendicular distances from the points to the line. Here, I used it as a dimension reduction device to evaluate the first two principal components for each observation vector and construct a scatter plot to check for multivariate normality, outliers, and so on.

In this contest, as a preliminary step of my thesis, I performed a Principal Component Analysis using the snpgdsPCA function of the SNPRelate Rpackage (Zheng, 2013).

### 3.3.2 Admixture analysis

To investigate genetic structure in the Candidate and Reference datasets, I used the model-based clustering algorithm implemented in the ADMIXTURE software (http://www.genetics.ucla.edu/software/admixture/, Alexander et al., 2009). It performs a maximum likelihood estimation of individual ancestries in the samples. Each analysis requires a hypothesized number of ancestral population (*K*) and assign individuals the ancestry proportion for each of them. The choice of the *K* value can prove difficult when the underlying population genetics of a species is poorly understood, as in our species. In order to identify the most supported number (*K*) of ancestral populations, I used the cross-validation procedure implemented in the tool. It is a model evaluation method that is better than simply looking at the residuals, although it represents the universally standard measure for model fit in linear models. Indeed, regression evaluation does not indicate how well a model can make new predictions on cases it has not already seen. One way to overcome this problem is to not use the entire data set; therefore, cross validation techniques tend to focus on not using the entire data set when building a model. In particular, in the K-fold cross validation procedure, which I applied in my thesis, partitions the non-missing genotypes into *K* roughly equally sized subsets (*folds*). For each fold it estimates a model on all the subsets except one, using the left out subset to test the model, by calculating a CV metric of choice. Finally, user receive the CV error as the average between the CV metric across subsets. As a result, the validation error gives an unbiased estimate of the predictive power of a model. For the

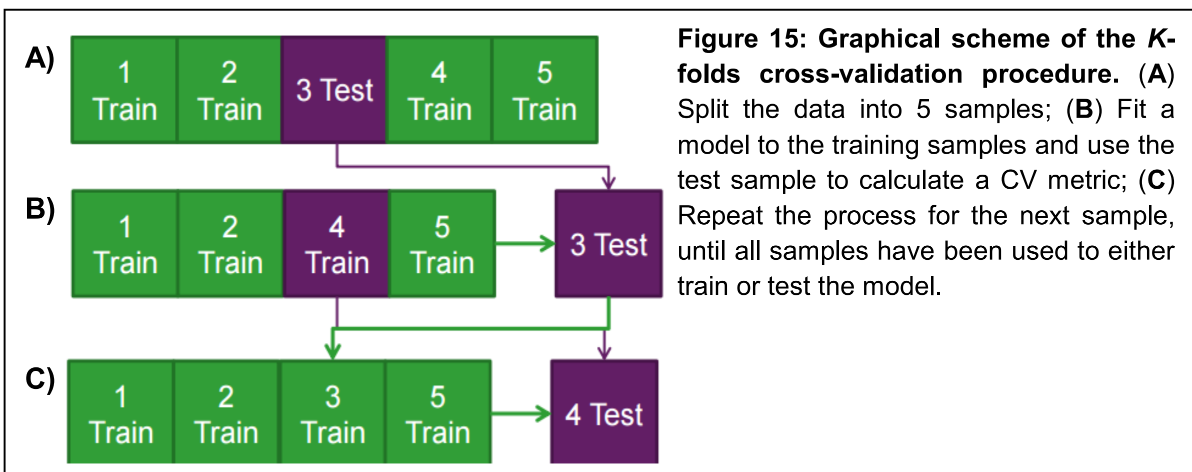



**Figure 15: Graphical scheme of the *K*-folds cross-validation procedure.** (**A**) Split the data into 5 samples; (**B**) Fit a model to the training samples and use the test sample to calculate a CV metric; (**C**) Repeat the process for the next sample, until all samples have been used to either train or test the model.

here presented analysis, I used the number of folds defined by default in ADMIXTURE software, that is equal to 5 (Figure 15)

I calculated the cross-validation error (CV) considering a number of *K* from 1 to 20. In both Reference and Candidate datasets 5 ancestral populations seem to best explain the genetic variation contained in the data. Then, I clustered individual genotypes exploring a range of *K* values between 2 and 12, and repeating the analysis 50 times for each *K* value.

The results were finally summarized and plotted using, respectively, the CLUMPP (http://www.stanford.edu/group/rosenberglab/clumpp.html; Jakobsson and Rosenberg, 2007) and Distruct software packages (Rosenberg et al., 2002; http://www.stanford.edu/group/rosenberglab/distruct.html).

### *3.4 Testing for the effects of natural selection*

Positive selection leaves a signature in the genome that we tried to detect in three ways.

In order to test for local adaptation at clock genes, I followed two parallel candidate-gene approaches. Firstly, I used BayeScan (http://cmpg.unibe.ch/software/BayeScan/; Foll and Gaggiotti 2008), a differentiation-based method assuming a Bayesian framework in which $F_{ST}$ is a model parameter. For each locus, and taking advantage of a reversible jump Markov Chain Monte Carlo algorithm, BayeScan estimates the scale of evidence in favour of a model including selection versus a model without selection. In particular, BayeScan scans for local adaptation in allele frequency data modelling separately a coefficient (ß) assuming an island model of demography and thus regarded as population-specific, and a locus-specific coefficient (α) that reflects the strength of selection acting on a particular locus. For each SNP, BayeScan estimates the posterior distribution under neutrality ($\alpha = 0$) and allowing for selection ($\alpha \neq 0$), thus computing the posterior odds ratio (*PO*) as an estimate of the support for the model of local adaptation compared to neutral demography. This way, if α is positive, the locus considered has potentially undergone positive selection, whereas, if α is negative, the locus is considered subjected to balancing selection. Outliers' detection levels depend on the estimation of the *PO*, which can be evaluated through the Jeffrey's scale of evidences (Jeffreys 1961). To control against false positive results, I calculated the expected value of $\log_{10}PO$ yielding a 1% of false discovery rate. BayeScan analyses have been performed on both the Candidate and the Reference datasets under identical run conditions. Specifically, these consisted of a prior odds of 1,000, meaning that I considered the selection model 1/1000 as likely as the neutral model for any given SNP, 20 pilot runs, 50,000 burn-in iterations, followed by 50,000 output iterations with a tinning interval of 10, resulting in 5,000 iterations for posterior estimation. We ran this analysis separately on the reference SNPs and

candidate SNPs under identical conditions. As suggested in the BayeScan manual, we removed from the analysis loci with very low minor allele frequencies (< 0.005), so that 14,707 SNPs of the Candidate dataset and 27,626 SNPs of the Reference dataset were in fact analysed.

We defined two significance threshold. The first one was a Model-based significance threshold, namely the expected value of the $\log_{10}$ of the posterior odd ratio (the support for the model of local adaptation relative to neutral demography) which would yield a 1% false-discovery rate based on the reference SNPs. The second threshold was empirical, hence called Reference-based significance threshold. This threshold is calculated on the Reference dataset and corresponds to upper confidence limit of the FST distribution in the Reference dataset; all SNPs in the Candidate dataset showing FST values beyond this limit were considered significant.

If the population structure does not correspond to Wright's island model, BayeScan may give biased results (Excoffier et al. 2009). To circumvent this problem, I turned to the second method, based on a hierarchical island model (Excoffier et al. 2009) and implemented in the Arlequin software (v3.5.1.2; Excoffier and Lisher 2010), which specifically takes into account co-ancestry among related subpopulations. Accordingly, I generated the expected neutral $F_{ST}$ distribution conditioned on heterozygosity using a subsample of 1,000 reference SNPs. To model the hierarchical genetic structure, in agreement with Excoffier and co-workers (2009), I defined 10 groups, each consisting of 100 demes. The expected distribution of $F_{ST}$ values for the reference SNPs was then generated simulating the effects of this structure for 50,000 replicates, by the Arlequin software. The same approach was then used to calculate $F_{ST}$ on the Reference dataset. Again, in agreement with the previous (BayeScan) analysis of local adaptation, I defined a Reference-based threshold, corresponding for the candidate SNPs to an $F_{ST}$ value greater than the 99% upper limit of the reference $F_{ST}$ distribution, and a Model-based threshold related to the simulated $F_{ST}$ distribution based on the 50,000 simulations.

In the third approach, I looked for local selection reflecting an environmental gradient. Indeed, loci involved in local adaptation can potentially be identified by an unusual correlation between allele frequencies and important ecological variables, or by extreme allele frequency differences between geographic regions. However, such comparisons are complicated by differences in sample sizes and the neutral correlation of allele frequencies across populations due to shared history and gene flow. Through a Bayesian approach, Coop and collaborators. (2010) developed BAYENV2 software (http://gcbias.org/bayenv/) which can overcome these difficulties. It can estimate the empirical pattern of covariance in allele frequencies between populations from a set of markers, and then uses this as a null model

for a test at individual SNPs (Coop et al., 2010). Compared to other tests based on pairwise or global $F_{ST}$, BAYENV can be used to identify SNPs with unusually large allele frequency differentiation, representing a powerful alternative.

Applying a Monte Carlo Markov Chain, I firstly computed the neutral covariance matrix based on 10,000 reference SNPs, thus summarizing the pattern of allelic frequency variance among the 26 populations, according to a simple drift model. This matrix of population differences (Ω) was then used to control for population history when testing for covariance between environmental variables and the population-specific allele frequencies at a given SNP. At each SNP, we then provide a measure of the support, a Bayes factor (BF), for a model where an environmental variable has a linear effect on the transformed allele frequencies compared to a model given by the covariance matrix alone. In other words, for each locus BAYENV2 computes the ratio of the posterior probability (*PO*) of the model of adaptation vs random drift, where the *PO* and the associated Bayes Factor (*BF*) represent the support for the model of local adaptation respect to a model of random drift. BAYENV2 allows also the estimation of the non-parametric Spearman's *ρ* statistic, which is less affected by extreme values. I analysed each SNP individually and determined the distribution of *PO*, *BF* and *ρ* separately for reference and candidate loci.

This parametric method assumes a linear effect of the environmental variable, and this can lead to spurious correlations in the presence of strong outliers. In order to identify candidate polymorphisms really showing a strong departure from null expectations, I defined two significance thresholds. Hereafter, I shall refer to the Single threshold when, for a given SNP, the Bayes Factor of the model of adaptation vs the null model is higher than 10; the Combined threshold means instead that a SNP has both BF>10 and a ρ-value higher than 0.25, as in Jaramillo-Correa et al (2015).

### *3.5 Effect of natural selection in the phenotype-associated dataset*

As a final step, I focused on the polymorphisms showing an association with altered or abnormal chronotypes and/or sleep disorders, at least in one populations. Due to divergent environments, local adaptation can lead differences in average phenotype between populations of the same species. Therefore, I verified if SNPs in the "Phenotype-associated" dataset are cases in which the covariance of their allele frequencies and ecological variants exceeds the expected covariance inferred from the neutral loci of the reference SNPs. Particularly, I wanted to evaluate the contribution or the effect of an environmental or 'continental' variable on the frequency of an allele at a SNP, while controlling for the correlation of allele frequencies across populations.

To reach this purpose, I researched the 15 SNPs of the "Phenotype-associated" dataset

among the results of the previously performed BAYENV2 analysis on the Candidate dataset. For each SNP I obtained a value of Bayes Factor and the non-parametric correlation coefficient.

# 4. RESULTS

## *4.1 Analysis of population structure*

### 4.1.1 Principal Component Analysis

We started from exploratory analyses of two genomic datasets, respectively the Candidate and the Reference datasets.

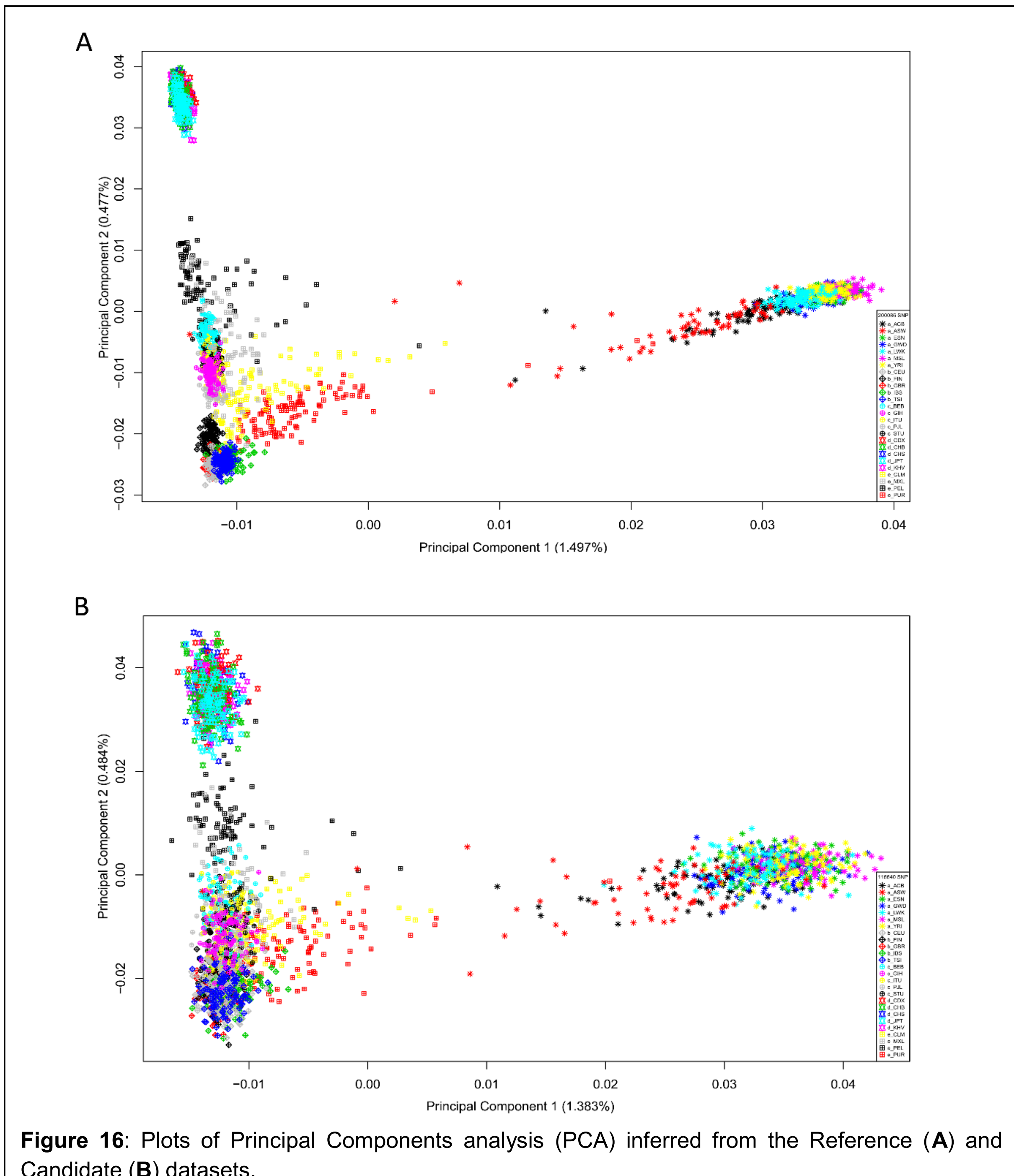



**Figure 16**: Plots of Principal Components analysis (PCA) inferred from the Reference (**A**) and Candidate (**B**) datasets.

The PCA plots of the Reference and Candidate SNPs show remarkable similarities (Figure 16A-B). In both cases, the African populations are spread along the first PC axis (accounting

in both cases for <1.5% of the total variance); admixed American populations occupy intermediate positions between Africa and, to the left, the Eurasian populations, which are distributed along the second PC axis (accounting in both cases for <0.5% of the total variance); Europeans are at the bottom and Asians at the top. At this level of resolution, it is impossible to judge whether the small differences observed between datasets are suggestive of specific selective processes affecting loci of the Candidate dataset; we can only say that these differences, if any, are not obvious.

### 4.1.2 Admixture analysis

Through ADMIXTURE software, I looked for the best clustering of genotypes of the Reference and Candidate datasets, in a number of clusters from *K*=2 to *K*=12. In agreement with previous studies (Rosenberg et al., 2002), at *K*=3, clusters correspond to the main continents (orange for Africa, blue for Europe, yellow for Asia), with the American populations showing, once again, evidence of admixture on a recognizable genomic background. The most likely number of clusters (minimum cross-validation error) was *K*=5 (Figure 17A-B).

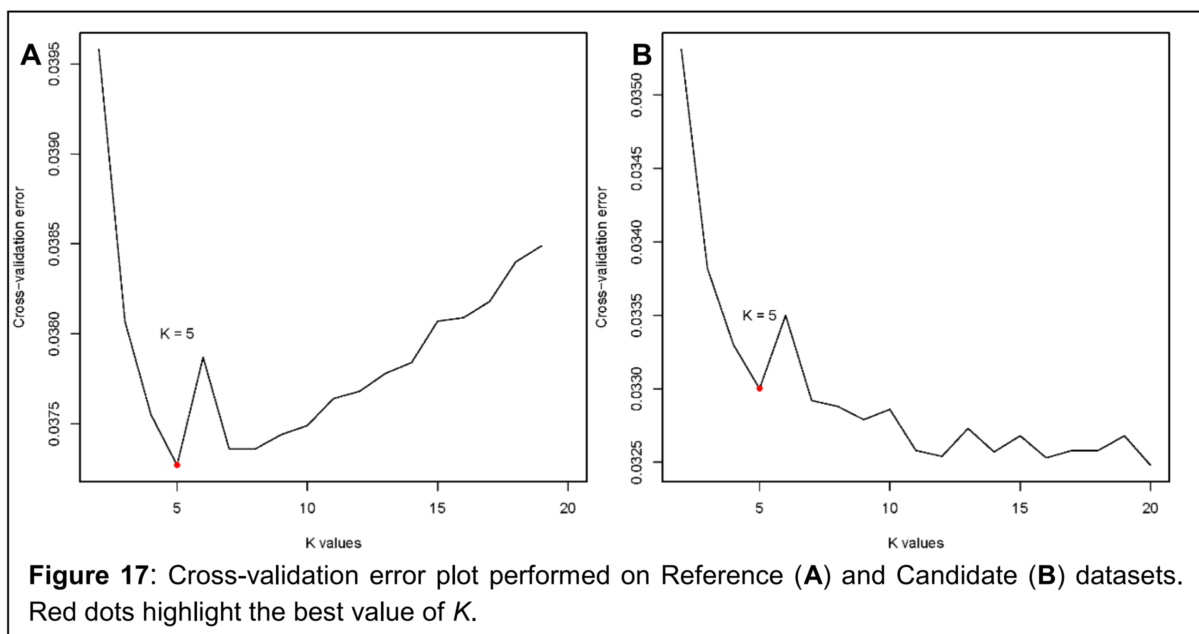
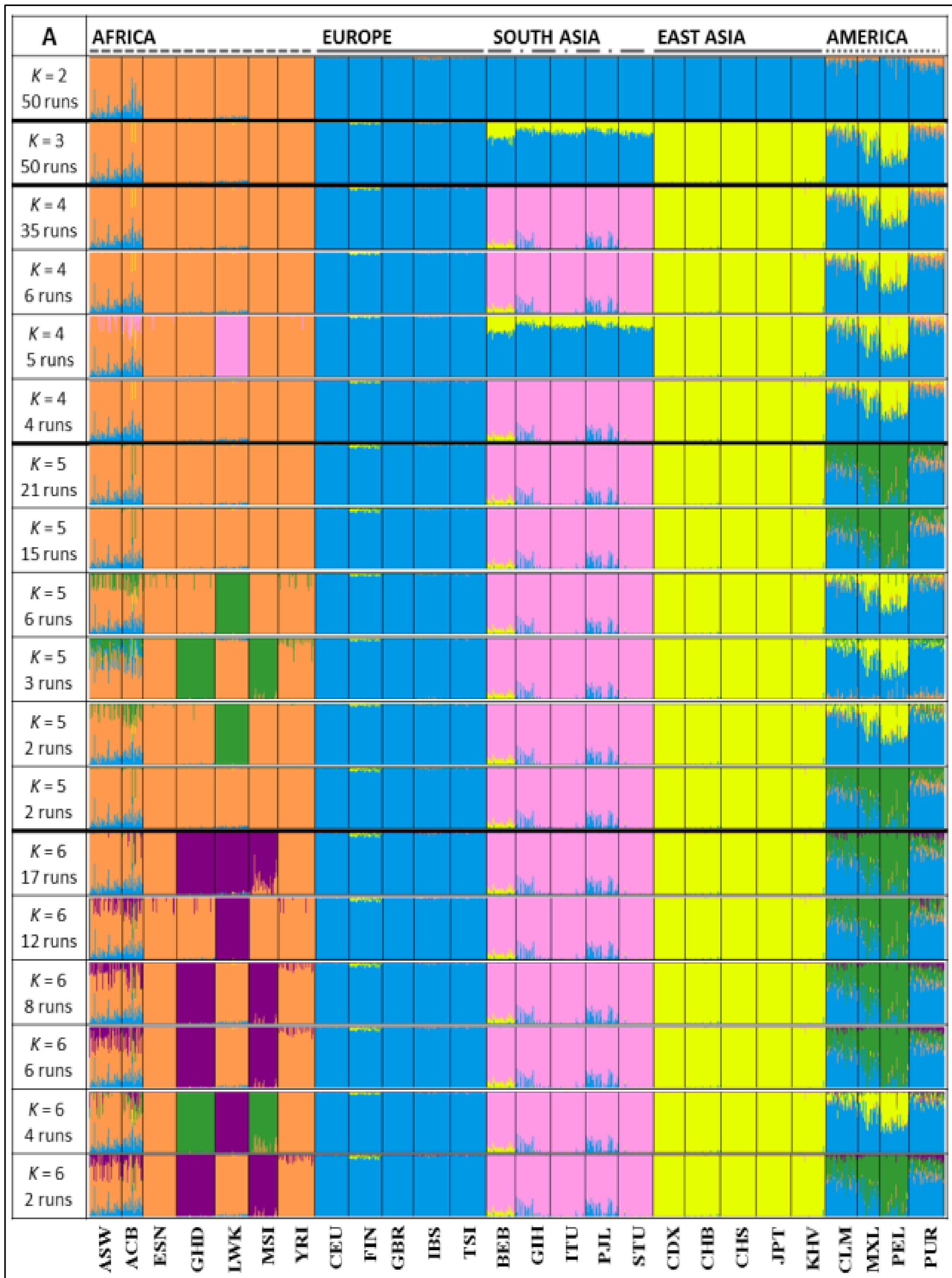



**Figure 17**: Cross-validation error plot performed on Reference (**A**) and Candidate (**B**) datasets. Red dots highlight the best value of *K*.

Here, two components were observed mainly in African and African-American individuals; because these components are present in all native African populations, it seems logical to consider them as an original African feature, rather than due to some kind of admixture. Europe, Southern Asia and Eastern Asia are well separated and internally homogeneous, whereas the American populations appear as very admixed, containing variable proportions of variants of African, European and Asian origin (Figure 18B). A remarkably similar pattern was observed for the Reference dataset (Figure 18A), which also showed the minimum cross-validation error at *K*=5. In short, these explorative analyses did not suggest that the

distribution of SNPs involved in the human circadian machinery is different from the distribution of random autosomal SNPs.



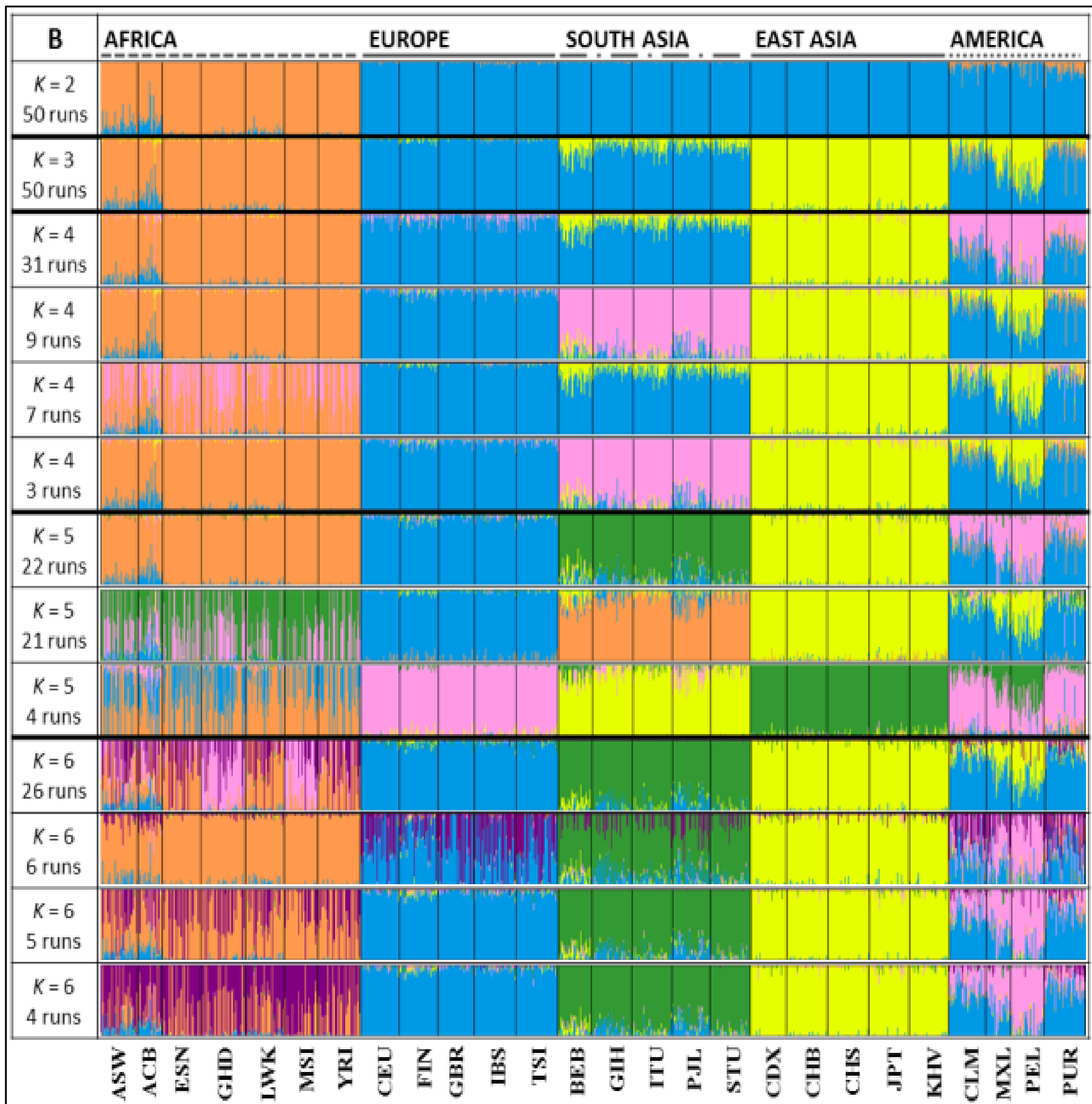



**Figure 18:** ADMIXTURE membership coefficients for *K* values from 2 to 6, performed on both the Reference (**A**) and the Candidate (**B**) datasets. For each value of *K*, 50 replications have been performed and, for some valued, I identified different clustering solutions. Here, I plotted different models supported at least by two replications. For each model, on the left there is indicated the *K* value and the number of runs. In every plot, different colour represents a cluster, and a single vertical line divided in *K* different colour, represents an individual. The black vertical lines divide individuals from different populations.

### *4.2 Testing for local adaptation*

BayeScan estimates of average population divergence were slightly higher for candidate SNPs than for the reference SNPs (0.125 vs 0.117). Therefore, the number of SNPs exceeded the 1% Model-based significance threshold is higher than the number of SNPs showing evidence of local adaptation using the Reference-based 1% threshold (Table1). Support for selection on candidate SNPs was very strong, with the 74% of SNPs having log10PO ratios > 3. This value, according to the Jeffrey's scale of evidence, is considered a

Decisive support for the selection model (Jeffreys, 1961). Consistent with this interpretation, the estimated locus-specific effect (α), was higher in the Candidate than in the Reference Dataset (0.32 vs 0.30, on average), reaching a mean value of 1.4 in the 1,250 positive selected polymorphisms of the Candidate Dataset.

**Figure 19:** Plot of the results of the Hierarchical Model analysis. Black points are the reference SNPs, grey ones are polymorphisms from the candidate dataset. A dotted grey line represents the 99% confidence interval based on the Reference dataset SNPs

Population divergence estimated through the Hierarchical Model identified 91 SNPs in 22 genes having extreme $F_{ST}$ values (higher than 0.3). As shown in Figure 19, this is a conservative choice, since 0.3 greatly exceeded the 99% confidence interval of the values simulated under a neutral model, represented by the dotted line (Figure 19). I found 1,073 SNPs in 25 genes with an $F_{ST}$ valued exceeding the 1% Reference-based threshold (Table 1), and38 of these SNPs having $F_{ST}$ values much greater than the highest $F_{ST}$ (0.37) of the Reference dataset. Taken together, the two approaches identified more polymorphisms with evidence of local adaptation considering the Reference-based threshold than the Model-based one (data are summarized in Table 1).

| **Significant threshold** | **Reference-based** | | **Model-based** | |
|---|---|---|---|---|
| **Approach** | BayeScan | Arlequin | BayeScan | Arlequin |
| **N° of SNPs** | 273 | 1 073 | 1 250 | 91 |
| **N° of genes** | 24 | 25 | 25 | 22 |
| **Shared SNPs** | *230 SNPs in 23 genes* | | *84 SNPs in 19 genes* | |

**Table 1**: Number of SNPs showing evidence of local adaptation, using both the Reference-based threshold (left) and the Model-based threshold (right). The number of SNPs giving significant results under both approaches are reported at the bottom.

Among these loci, the highest percentages of SNPs with signatures of local adaptation are located in the RORA gene (18% and 25% for the Reference-based and Model-based thresholds, respectively; Table 2), a key circadian transcriptional activator required for normal Bmal1 expression (Sato et al., 2004).

| **Reference-based** | | **Model-based** | |
|---|---|---|---|
| **Gene** | **% of SNPs** | **Gene** | **% of SNPs** |
| RORA | 18 | RORA | 25 |
| NPAS2 | 11 | CKIδ | 19 |
| BMAL1 | 10 | AVPR1B | 12 |
| | | TIM | 11 |

**Table 2**: Genes showing at least 10% of SNPs with evidence of local adaptation.

To explore the possibility that the distributions of allelic variants at clock genes might simply reflect one or a few environmental factors, we then looked for significant associations between allele frequencies and eight environmental variables (geographic distance, photoperiod, humidity, ratio between longwave and shortwave radiation fluxes, precipitation, temperature, latitude and longitude) by means of the software BAYENV2 (Gunther and Coop 2013). This way, we identified 190 SNPs in 23 genes under the Single threshold, and 69 SNPs in 20 genes under the Combined threshold; all these SNP frequencies correlate with, at least, one of the environmental variables (Table 3).

| **Significant threshold** | **Single** | **Combined** |
|---|---|---|
| **N° of SNPs** | 190 | 69 |
| **N° of genes** | 23 | 20 |

**Table 3**: Number of SNPs showing evidence of correlation with environmental variables, using both the "Single" and "Combined" thresholds in the BayEnv analysis.

Among the environmental factors studied here, temperature and precipitation show the highest percentage of correlations with allele frequencies, respectively 21% and 12% of all significant correlations when considering the Single threshold, and 25% and 17% considering the Combined threshold.

Considered all together, the overlap among the results of the three analyses was limited; we found only 2 SNPs exhibiting both elevated population differentiation and covariance with one or more environmental variables (Table 4). This is not unexpected, since the tests based on correlations would identify only clinal spatial patterns, whereas an excess variance does not necessarily reflect any spatial variables.

| Significant threshold | Gene | SNP | Correlated environmental variables |
|---|---|---|---|
| Reference-based | NPAS2 | rs72976842 | Latitude<br>Ratio L/S wave radiation fluxes |
| Model-based | AANAT | rs4647868 | Ratio L/S wave radiation fluxes |

**Table 4**: Two SNPs showing both evidence of local adaptation and significant correlation with, at least, one environmental factor. For each SNP the significance threshold and the correlated environmental variable is reported.

#### *4.3 Effect of natural selection in the phenotype-associated dataset*

Finally, we wanted to understand whether SNPs in the “Phenotype-associated” dataset have been possible targets of adaptation phenomena. None of our Phenotype-Associated SNPs showed significant departures from null expectations in the previous analyses based on extreme $F_{ST}$ values (i.e. BayeScan and Hierarchical Model). However, when considering candidate SNPs with BF higher than 99% of the reference values in the BAYENV2 analysis, we could identify five polymorphisms showing significant correlation with at least one of the environmental factors here analysed. These polymorphisms, falling in five crucial clock genes, along with the correlated environmental variables, are detailed in Table 5.

| Gene | SNP | Correlated environmental variable(s) | Chronotypes associated | References |
|---|---|---|---|---|
| NPAS2 | *rs7598826* | Distance from Addis Abeba<br>Ratio L/S wave radiation fluxes<br>Mean longwave radiation flux | Alteration of sleep duration | Allebrandt et al., 2010 |
| PER2 | *rs934945* | Mean Humidity Rate | Diurnal preferences | Lee et al., 2011 |
| PER3 | *rs228730* | Mean Precipitation Rate | DSPS | Archer et al., 2010 |
| CKId | *rs7209167* | Photoperiod | Alteration of sleep duration | Allebrandt et al., 2010 |
| OPN4 | *rs2675703* | Distance from Addis Abeba<br>Longitude | Variation in sleep onset and chronotypes | Roeckelin et al., 2012 |

**Table 5**: Five SNPs in the phenotype-associated dataset showing significant correlation with at least one environmental variable. For each polymorphism, we reported the abnormal chronotypes or the sleep disorder to which the polymorphism has been previously associated.

## 5. DISCUSSION

Taken together, the results of my analyses do show support for the view whereby various components of the human circadian clock evolved under the effects of positive selection, but also that selective pressures were rather weak upon the individual loci.

Previous studies in human populations investigated the association of the circadian clock genes with pathological conditions, such as metabolic and affective disorders. For example, Forni et al. (2014) analyzed whole genome polymorphisms data in populations of the CEPH panel to investigate how the seasonal photoperiod variation might have exerted a selective pressure during the expansion of modern humans out of Africa. They suggested that this expansion influenced adaptive evolution at circadian regulatory loci and at risk variants for psychiatric and neurologic diseases. Lane et al. (2016) performed a genome-wide association study of self-reported chronotypes of more than 100 000 UK individuals, identifying 12 variants in genes involved in the circadian clock mechanism and in the light-sensing pathways (possibly related to schizophrenia, educational attainment and BMI) . A similar study has been conducted by Hu et al (2016), who searched for correlations between whole genome variants and a self-reported morningness in the 23andMe participant cohort, finding significant signals in circadian and phototransduction pathways.

My thesis survey represents a broader effort to explore specifically human clock genes' variation at the worldwide level and using polymorphisms data coming from complete genomes; this way, we could draw inferences on their evolution and on the effect of potentially relevant factors in the environment. The Candidate and the Phenotype-associated datasets here analysed include SNPs associated either generally with the circadian clock, or specifically to chronotypes alterations. I did not focus on risk alleles associated with other diseases such as schizophrenia, bipolar disorder, and restless leg syndrome, where potential signatures of natural selection were recently detected (Forni et al. 2014).

Inferring natural selection by means of PCA or Cluster analysis is difficult or impossible, and in fact it was not among the purposes of these preliminary analyses Rather, I tried to identify whether there have been processes acting on the whole pathway that can have modified the genetic structure of clock and clock-related genes respect to that present in neutral loci. The worldwide genomic structure of genes related to circadian rhythms was so far unknown, so the analysis here performed represents its first description. The initial analyses of population structure show that the distribution of the SNPs of the Candidate dataset, including clock, clock-controlled and clock-related genes, does not depart from that observed in the analysis of neutral genome diversity (see also Rosenberg et al. 2002; Li et

al. 2008). Although such a result does not mean that selection was irrelevant, certainly it does not point to simple selective mechanisms affecting single genes as major determinants of population differences.

Inferring positive selection from genomic data is complicated by the fact that patterns in the data may indeed reflect selection, but also the effects of population history, or a combination of both. For some sixty years now, the presence of outliers in the distribution of genetic variances, i.e. of loci showing either very low or very high levels of variation, has been taken as suggestive of phenomena affecting single loci (such as selection), as opposed to the whole genome (such as genetic drift and gene flow) (Cavalli-Sforza, 1966). The availability of large amounts of information about non-genic regions has increased the power of tests exploiting this basic idea (Novembre and Di Rienzo, 2009).

Because the Admixture and PCA analyses did not point to obvious differences between clock genes and presumably neutral genome regions, we chose to combine Because the selection signal was not apparent in our datasets data when jointly analyzed, there seemed to be no alternatives to combining different approaches to detect the signature of selection at individual loci. (see e.g. Keller et al. 2012; Seeb et al. 2014). Whether or not the island model assumed by BayeScan can faithfully represent migration relationships at this scale is open to discussion. In principle, a hierarchical island model seems to be more realistic and more conservative than a standard island model, but in practice there is no way to tell. Thus I decided to combine the two approaches, in order to identify a (smaller) subset of SNPs identified by both methods, and hence to have a higher chance to tell natural selection from the background of demographic history.

The results of these analyses point to several SNPs departing from null expectations in their distribution, some of them very significantly. Overall, I identified 230 and 84 SNPs of clock genes with evidence of local adaptation, considering the Reference-based and the Model-base thresholds, respectively. Most such polymorphisms were in the BMAL1, NAPS2, TIM, RORA, and CKIδ genes, suggesting selection action might have affected all circadian clocks components (positive, negative and clock-related genes).

In addition, and contrary to what was observed in one (Cruciani et al. 2008) but not all (Forni et al. 2014) previous studies on human clock genes, I also found some evidence of correlation between circadian clock SNPs and environmental factors.

Two SNPs emerge from the joint analyses of local and clinal adaptation, namely *rs72976842* in NPAS2, and *rs4647868* in AANAT. NPAS2 (Neuronal PAS domain protein 2) is a transcription factor paralog of CLOCK expressed mainly in the mammalian forebrain. It has been shown in mice that both clock proteins can independently heterodimerize with BMAL1 in the SCN to maintain molecular and behavioural rhythmicity (Reick et al., 2001). An overlapping role of CLOCK and NPAS2 has been also showed in the circadian

expression of liver FVII, a key liver protease (Bertolucci et al., 2008). In humans, epidemiologic evidence linked NPAS2 with a variety of disorders, including winter depression (Partonen et al., 2008). Furthermore, a NPAS2 missense mutation (*rs2305160*) is associated with risk of cancer (Zhu et al., 2008), but that SNP is not in linkage disequilibrium with the one I found to be positively selected, namely *rs72976842* ($r^2$=0.014).

The second locus is located on AANAT (Arylalkylamine N-acetyltransferase), a gene which codes for a key enzyme involved in daily melatonin synthesis and its transcription is regulated by the circadian clock via the E-box promoter elements. In AANAT, *rs28936679* (Table 3) defines an amino acid substitution (Ala129Thr), associated (albeit in a sample of only 50 Japanese individuals) with delayed sleep phase syndrome (DSPS), a disorder characterized by a significantly delayed sleep onset and wake-up times (Hohjoh et al., 2003). At this locus I did not find any evidence of correlation with environmental factors. In both cases, it remains to be investigated why selection may have favoured the spread of allelic variants with such negative effects.

Evidence of correlation between ecological factors and 5 SNPs from the Phenotype-associated dataset suggest that the genetic variation associated to some crucial components of the circadian clock is variously influenced by environmental factors (Table 5). Three such polymorphisms are in clock genes (*rs7598826* in NPAS2, *rs934945* in Per2 and, *rs228730* in Per3) and 2 in clock-related genes CKIδ (*rs7209167*) and OPN4 (*rs2675703*).

The polymorphism *rs7598826*, located on NPAS2, resulted to be the one with the highest number of correlated environmental factor: the geographic distances, and two factors including the radiation fluxes. From a previous study, it is known to correlate with sleep duration (Allebrandt et al. 2010).

The two SNPs in PERs correlate with humidity and precipitation mean rates and have been associated, respectively, with diurnal preferences (*Per2*, *rs934945*: Lee at al., 2011; Per3, *rs228697*: Hida et al. 2014) and DSPS (*Per3*, *rs228730*: Archer et al., 2010). Both *Per2* and *Per3* participate in timekeeping mechanism in the central (SCN) and peripheral (pituitary gland and lung) oscillators, respectively (Pendergast, et al. 2010). A recent study demonstrated that the phosphorylation of the PER2 protein is the mechanism underpinning the "temperature compensation", a key feature of the circadian clock that allows to the organisms to maintain a 24-hours period regardless of the environmental temperature (Zhou et al, 2015). A human *Per2* mutation (S662G, *rs121908635*) was well investigated because it is responsible of the Familial Advanced Sleep Phase Syndrome (FASPS), a rare phenotype characterized by early sleep times and early-morning awakening (Toh et al., 2001). Consistently with the hypothesis that circadian clock system is a complex network of genes, evidences of associations between loci and chronotypes have been found also not

only in the clock genes, but also in near regions and genes. In fact, recently, a polymorphism (*rs55694368*) located 120 kb upstream of PER2 have been shown to be strongly associated with self-reported morningness individuals. This SNP is located in a DNAse hypersensitive site (HDS) for five cell types, including pancreas adenocarcinoma, B-lymphocite, medulloblastoma and CD4+ cells. Interestingly, the SNP has been shown to be associated with the human familial ASPS (Hu et al., 2016). In addition, Lane and collaborators (2016) reported *rs35333999* in PER2 be suggestive that common variants contribute to variation in chronotypes

Also Per3 has been deeply studied in detail because its involvement with both delayed sleep phase syndrome and extreme diurnal preferences (Archer et al., 2003). In addition, the interest for this gene is due to the fact that some 10% of the human population is homozygous for the 5-repeat allele of a VNTR polymorphism within it, and these people show morning preferences (Archer et al., 2003). By contrast, the homozygous genotype for the 4-repeat allele has been associated with evening preferences (Archer et al., 2003). In addition, a suggestive signal of correlation of a locus (*rs7545893*) located ~30kb upstream of the untyped VNTR region with chronotypes has recently been found in a large UK sample (Lane et al., 2016). We did not investigate this VNTR polymorphism, but another SNP in Per3 (rs228697), previously associated with diurnal sleep preference and free-running type (Hida et al. 2014), was eliminated from our analysis because it is in strong linkage disequilibrium (r2=0.935) with rs228730. Work is in progress to investigate the interaction between the Per3 VNTR and SNP variants. At 8 kb downstream the PER3 gene, the polymorphism *rs11121022* shows strongly evidences of association with morningness (Hu et al., 2016). It is involved in the regulation of three regulatory motifs and it cat affect the sensitivity of the circadian system to light; therefore, it is involved in the sleep/wake activity (Hu et al., 2016).

As for the CKIδ gene, it controls nuclear transport and degradation of core elements of the circadian clock such as PER1 and PER2 and determines the circadian period length through the regulation of the speed and rhythmicity of their phosphorylation (Partch et al. 2014). It is intriguing to note that *rs7209167* in CKIδ shows a significant correlation with photoperiod and has been associated with the alteration of sleep duration (Allebrandt et al., 2010). CKIδ is currently the target of pharmacological studies aimed at shifting or resetting the phase of circadian rhythm for the treatment of circadian rhythm disorder, such as the FASPS determined by a missense mutation (T44A, *rs104894561*) (Xu et al., 2005). *rs104894561* was not included in our analysis because population data are not available in 1000Genome browser.

Finally, OPN4, also called Melanopsin, is a circadian photopigment expressed within the ganglion and amacrine cell layers of the mammalian retina (Lucas, 2013), mediating non-

image-forming responses to light, such as the sleep induction, the pupillary light response, and others. The mouse Melanopsin gene is present in two fully functional isoforms: the short (OPN4S) and the long (OPN4L) isoform. Recent papers showed that these different isoforms of OPN4 mediate different behavioural responses to light (Jagannath et al., 2015). Both isoforms are also present in the human genome, but the functional results require confirmation. *rs2675703* in OPN4, the SNP that we found correlated with two geographical parameters, has been previously associated with variation in sleep onset and chronotypes (Roeckelin at al., 2012). Recently, activator or inhibitor of Melanopsin have been proposed as treatments to mimic light and darkness to manage sleep disturbances and mood disorder in normal and blind patients (Hatori and Panda, 2010).

Although all these findings make biological sense, and may have implications for the development of drugs against sleep disturbances and disorders, very few SNPs of this study correlate with some of the main environmental variables. I interpret these findings as suggestive that mechanisms underpinning the evolution of human circadian clock, much like those related with other complex traits, are the result of more complex interactions, ones which can hardly be discovered focusing only on SNPs frequencies.

Investigations on how circadian systems are adapted to different latitudes in humans are at an early stage. Evidence of genetic differences likely due to selection caused by environmental conditions along latitudinal clines are mainly available for insects and fishes (Kyriacou et al., 2008; O'Malley and Banks, 2008). Analysis conducted on Drosophila melanogaster (Kyriacou et al., 2008) showed that latitude and photoperiod could indeed affect the evolution of circadian clock genes.

The selection mechanisms we could investigate at the genomic level are clearly unlikely to affect all clock genes, and even less so to act with the same intensity on most of them. The emerging picture seems one in which local selective pressures had an effect on one or a few components of the human biological clock. However, these effects are likely minor, as shown by the absence of significant differences in the global patterns of variation for clock and clock-related genes on the one hand, and the set of presumably neutral markers I analysed on the other.

## 6. FUTURE PERSPECTIVES

Candidate gene association studies have reported variation associated with morningness or evening preferences in the clock genes. However, these studies have often had limited reproducibility, mainly due to the small sample sizes, the heterogeneity in chronotypes assessment and inadequate correction for population structure. This highlights the necessity of replication in larger samples before firm conclusions can be drawn. Thanks to the technological developments in high-throughput sequencing, today there is an increasing number of available whole genome sequences from worldwide human populations. Therefore, in order to validate the results so far obtained, in near future more populations could be included in the study. Particularly, it could be interesting the involvement of more samples living in the southern hemisphere. Indeed, a larger sample will certainly help telling apart the effects of positive selection from those of shared demographic history. It will be crucial to standardize the criteria for assessing phenotypes. The questionnaires currently used have the advantage to be easy to fill, but leave ample space to subjective considerations. Future analyses defining phenotypes on the basis of biochemical or clinical tests may pave the ground for a higher reproducibility of the results, and hence for a deeper understanding of the effects clock-gene diversity.

In parallel, and due to the thigh interaction between gene and environment, analyses could be replicated including more ecological variables. Indeed, the worldwide distribution of our specie lead to different environmental triggering factors that could have affect our evolution. Among that, it could be interesting evaluate investigate the role of alimentary habits characterizing different populations.

In the last years, many advances have been made also in the field of ancient genomes. Thus, thanks to the massive sequence throughput of next generation sequencing platforms, an increasing number of archaic genome sequences are available. The direct investigation of the diversity of ancient populations could help shedding light on the evolutionary processes, such as ones that played a role in the evolution of genes involved in human circadian rhythms. Therefore, we would compare the modern genetic variability at clock genes with the archaic one, including Neandertal, Denisovan, Ust'-Ishim ones and genomes from upper Palaeolithic and late Pleistocene (Sarkissian et al., 2015).

However, an even more crucial leap forward will be represented by the development of new methods allowing one to jointly consider several genes and their interactions, and compare them with a large set of variables in the environment.

While population genetic studies, as the one here presented, and biochemical approaches largely contributed to uncover the molecular bases of circadian rhythmicity, the biochemistry is investigating how the circadian clock is modulated. Strategies such as targeting the circadian oscillator by small chemical compounds are increasingly developing. The importance of a robust circadian timing for health is increasingly recognized and many studies are focusing on the identification of small molecules capable of modulating circadian clocks (Wallach and Kramer, 2015). Therefore, the results of genetic approaches, as the ones implemented in my thesis, could be important for the further characterize the molecular basis of circadian rhythm generation. Moreover, they will also allow us to develop novel treatment strategies for alteration and/or diseases that are associated with altered or disrupted clocks.

## 7. LITERATURE CITED

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA et al. (2015). A global reference for human genetic variation. Nature 526(7571):68-74.

Abdi H, Williams LJ. (2010). Principal component analysis. WIREs Comp Stat, 433–459.

Alexander DH, Novembre J, Lange K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome research 19, 1655-1664.

Allebrandt KV, Roenneberg T. (2008). The search for circadian clock components in humans: new perspectives for association studies. Braz J Med Biol Res. 41(8):716-21. Review.

Allebrandt KV, Teder-Laving M, Akyol M, Pichler I, Müller-Myhsok B, Pramstaller P, Merrow M, Meitinger T, Metspalu A, Roenneberg T. (2010). CLOCK gene variants associate with sleep duration in two independent populations. Biol Psychiatry 67(11):1040-1047.

Arbiza L, Zhong E, Keinan A. (2012). NRE: a tool for exploring neutral loci in the human genome. BMC Bioinformatics 14(13):301.

Archer SN, Robilliard DL, Skene DJ, Smits M, Williams A, Arendt J, von Schantz M. (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.

Archer SN, Viola AU, Kyriakopoulou V, von Schantz M, Dijk DJ. (2008). Inter-individual differences in habitual sleep timing and entrained phase of endogenous circadian rhythms of BMAL1, PER2 and PER3 mRNA in human leukocytes. Sleep 31(5):608-617.

Archer SN, Carpen JD, Gibson M, Lim GH, Johnston JD, Skene DJ, von Schantz M. (2010). Polymorphism in the PER3 promoter associates with diurnal preference and delayed sleep phase disorder. Sleep 33(5):695-701.

Aschoff J. (1960). Exogenous and endogenous components in circadian rhythms. Cold Spring Harbor Symposia on Quantitative Biology: Volume XXV. Biological Clocks. New York: Cold Spring Harbor Press. pp. 11-28.

Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB. (2003). Human population genetic structure and inference of group membership. Am J Hum Genet. 72(3):578-89.

Bailey SL, Heitkemper MM. (2001). Circadian rhythmicity of cortisol and body temperature: morningness-eveningness effects. Chronobiol Int. 18(2):249-61.

Barbosa AA, Pedrazzoli M, Koike BD, Tufik S. (2010). Do Caucasian and Asian clocks tick differently? Braz J Med Biol Res. 43(1):96-9.

Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL. (1997). An apportionment of human DNA diversity. Proc Natl Acad Sci U S A. 94(9):4516-9.

Bedont JL, Blackshaw S. (2015). Constructing the suprachiasmatic nucleus: a watchmaker's perspective on the central clockworks. Front Syst Neurosci. 9:74.

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN. (2004). Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet. 74(6):1111-1120.

Bertolucci C, Cavallari N, Colognesi I, Aguzzi J, Chen Z, Caruso P, Foá A, Tosini G, Bernardi F, Pinotti M. (2008). Evidence for an overlapping role of CLOCK and NPAS2 transcription factors in liver circadian oscillators. Mol Cell Biol. 28(9):3070-5.

Bollinger T, Schibler U. (2014). Circadian rhythms - from genes to physiology and disease. Swiss Med Wkly. 144:w13984.

(book) Brown FA Jr, Hastings JW, Palmer JD. (1970). The Biological Clock - Two Views. New York: Academic Press.

Buhr ED, Yoo SH, Takahashi JS. (2010). Temperature as a universal resetting cue for mammalian circadian oscillators. Science. 330(6002):379-85.

Buhr ED, Takahashi JS. (2013). Molecular components of the Mammalian circadian clock.Handb Exp Pharmacol. (217):3-27. Review.

Busino L, Bassermann F, Maiolica A, Lee C, Nolan PM, Godinho SI, Draetta GF, Pagano M. (2007). SCFFbxl3 controls the oscillation of the circadian clock by directing the degradation of cryptochrome proteins. Science. 316(5826):900-4.

Carpen JD, Archer SN, Skene DJ, Smits M, von Schantz M. (2005). A single-nucleotide polymorphism in the 5'-untranslated region of the hPER2 gene is associated with diurnal preference. J Sleep Res. 14(3):293-7.

Cavalli-Sforza LL. 1966. Population structure and human evolution. Proc R Soc. 164:362-379.

Chang AM, Buch AM, Bradstreet DS, Klements DJ, Duffy JF. (2011). Human diurnal preference and circadian rhythmicity are not associated with the CLOCK 3111C/T gene polymorphism. J Biol Rhythms. 26(3):276-9.

Chatfield C, Collins AJ. (1980). Introduction to Multivariate Analysis. Chapman & Hall, London and New York

Chong SY, Ptáček LJ, Fu YH. (2012). Genetic insights on sleep schedules: this time, it's PERsonal. Trends Genet. 28(12):598-605. Epub 2012 Aug 28. Review.

Ciarleglio CM, Ryckman KK, Servick SV, Hida A, Robbins S, Wells N, Hicks J, Larson SA, Wiedermann JP, Carver K et al. (2008). Genetic differences in human circadian clock genes among worldwide populations. J Biol Rhythms 23(4):330-340.

Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, Absher D, Myers RM, Cavalli-Sforza LL, Feldman MW, Pritchard JK. (2009). The role of geography in human adaptation. PLoS Genet. 5(6):e1000500.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK. (2010). Using environmental correlations to identify loci underlying local adaptation. Genetics 185(4):1411-1423.

Cruciani F, Trombetta B, Labuda D, Modiano D, Torroni A, Costa R, Scozzari R. (2008). Genetic diversity patterns at the human clock gene Period 2 are suggestive of population-specific positive selection. European Journal of Human Genetics 16(12):1526-1534.

Dauvilliers Y, Maret S, Tafti M. (2005). Genetics of normal and pathological sleep in humans. Sleep Med Rev. 9(2):91-100. Review.

De Mairan JJdO. (1729). “Observation botanique“. Histoire de l’Academie Royale des Sciences, 35-36.

Der Sarkissian C, Allentoft ME, Ávila-Arcos MC, Barnett R, Campos PF, Cappellini E, Ermini L, Fernández R, da Fonseca R, Ginolhac A et al. (2015). Ancient genomics. Philos Trans R Soc Lond B Biol Sci. 370(1660):20130387.

DiTacchio L, Le HD, Vollmers C, Hatori M, Witcher M, Secombe J, Panda S. (2011). Histone lysine demethylase JARID1a activates CLOCK-BMAL1 and influences the circadian clock. Science. 333(6051):1881-5.

Do MT, Yau KW. (2010). Intrinsically photosensitive retinal ganglion cells. Physiol Rev. 90(4):1547-81.

Doi M, Hirayama J, Sassone-Corsi P. (2006). Circadian regulator CLOCK is a histone acetyltransferase. Cell. 125(3):497-508.

Duffy JF, Rimmer DW, Czeisler CA. (2001). Association of intrinsic circadian period with morningness-eveningness, usual wake time, and circadian phase. Behav Neurosci. 115(4):895-9.

Dunlap JC. (1999). Molecular bases for circadian clocks. Cell. 96(2):271-90. Review.

Duong HA, Robles MS, Knutti D, Weitz CJ. (2011). A molecular mechanism for circadian clock negative feedback. Science. 332(6036):1436-9.

Ebisawa T, Uchiyama M, Kajimura N, Mishima K, Kamei Y, Katoh M, Watanabe T, Sekimoto M, Shibui K, Kim K, et al. (2001). Association of structural polymorphisms in the human period3 gene with delayed sleep phase syndrome. EMBO Rep. 2(4):342-6.

Edery I, Rutila JE, Rosbash M. (1994). Phase shifting of the circadian clock by induction of the *Drosophila* period protein. Science. 263(5144):237-40.

Edwards KD, Lynn JR, Gyula P, Nagy F, Millar AJ. (2005). Natural Allelic Variation in the Temperature-Compensation Mechanisms of the Arabidopsis thaliana Circadian Clock.Genetics.170(1):387-400. doi:10.1534/genetics.104.035238.

Engelen E, Janssens RC, Yagita K, Smits VA, van der Horst GT, Tamanini F. (2013). Mammalian TIMELESS is involved in period determination and DNA damage-dependent phase advancing of the circadian clock. PLoS One. 8(2):e56623.

Excoffier L, Hofer T, Foll M. (2009). Detecting loci under selection in a hierarchically structured population. Heredity (Edinb) 103(4):285-298.

Excoffier L, Lischer HEL. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564-567.

Fabbri M, Antonietti A, Giorgetti M, Tonetti L, Natale V. (2007). Circadian typology and style of thinking differences. Learning and Individual Differences 17:175-180.

Fisher RA. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. Transactions of the Royal Society of Edinburgh, 52, pp 399-433.

Fisher SP, Foster RG, Peirson SN. (2013). The circadian control of sleep. Handb Exp Pharmacol. 217:157-183.

Foll M, Gaggiotti O. (2008). A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977-993.

Forni D, Pozzoli U, Cagliani R, Tresoldi C, Menozzi G, Riva S, Guerini FR, Comi GP, Bolognesi E, Bresolin N, et al. (2014). Genetic adaptation of the human circadian clock to day-length latitudinal variations and relevance for affective disorders. Genome Biol 15(10):499.

Foster RG, Roenneberg T. (2008). Human responses to the geophysical daily, annual and lunar cycles. Curr Biol. 18(17):R784-R794.

(book) Frederock MB. (2015). “How the Earth was formed”, Max B. Frederick, Publishing.

Gachon F, Olela FF, Schaad O, Descombes P, Schibler U. (2006). The circadian PAR-domain basic leucine zipper transcription factors DBP, TEF, and HLF modulate basal and inducible xenobiotic detoxification. Cell Metab. 4(1):25-36.

Gachon F, Leuenberger N, Claudel T, Gos P, Jouffe C, Fleury Olela F, de Mollerat du Jeu X, Wahli W, Schibler U. (2011). Proline- and acidic amino acid-rich basic leucine zipper proteins modulate peroxisome proliferator-activated receptor alpha (PPARalpha) activity. Proc Natl Acad Sci U S A. 108(12):4794-9.

Gallego M, Virshup DM. (2007). Post-translational modifications regulate the ticking of the circadian clock. Nat Rev Mol Cell Biol. 8(2):139-48. Review.

Gareau JR, Lima CD. (2010). The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. Nat Rev Mol Cell Biol. 11(12):861-71. Review.

Gaspar L, Brown SA. (2015). Measuring circadian clock function in human cells. Methods Enzymol. 552:231-56. Review.

Germain A, Kupfer DJ. (2008). Circadian rhythm disturbances in depression. Hum Psychopharmacol. 23(7):571-85. doi: 10.1002/hup.964. Review.

Gibertini M, Graham C, Cook MR. Self-report of circadian type reflects the phase of the melatonin rhythm. Biol Psychol. (1999)50(1):19-33.

Groth P, Weiss B. (2006). Phenotype Data: A Neglected Resource in Biomedical Research? Current Bioinformatics 1:347-358.

(book) Gillespie JH. (2010). Population genetics: a concise guide. JHU Press.

Godinho SI, Maywood ES, Shaw L, Tucci V, Barnard AR, Busino L, Pagano M, Kendall R, Quwailid MM, Romero MR, et al. (2007). The after-hours mutant reveals a role for Fbxl3 in determining mammalian circadian period. Science. 316(5826):897-900.

Guernier V, Hochberg ME, Guégan JF. (2004). Ecology drives the worldwide distribution of human diseases. PLoS Biol. 2(6):e141.

Gunther T, Coop G. 2013. Robust identification of local adaptation from allele frequencies. Genetics 195(1):205-220.

Haldane JBS. (1949). Disease and evolution. Ric. Sci. Suppl. A 19: 68–76 [often requoted, e.g., in Dronamraju, K. (Editor), 1990, Selected Genetic Papers of J.B.S. Haldane, Garland Publishing, New York/London].

Hamblin MT, Di Rienzo A. (2000). Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am J Hum Genet. 66(5):1669-1679.

Hardy, G. H. (2003). Mendelian proportions in a mixed population. 1908. The Yale Journal of Biology and Medicine, 76(2), 79–80.

Harmar AJ. (2003). An essential role for peptidergic signalling in the control of circadian rhythms in the suprachiasmatic nuclei. J Neuroendocrinol. 15(4):335-8. Review.

Harris EE, Meyer D. (2006). The molecular signature of selection underlying human adaptations. Am J Phys Anthropol. 49:89-130.

Hastings MH, Maywood ES, Reddy AB. (2008). Two decades of circadian time. J Neuroendocrinol. 20(6):812-9.

Hatori M, Panda S. (2010). The emerging roles of melanopsin in behavioral adaptation to light. Trends Mol Med. 16(10):435–446.

Hellenthal G, Busby GB, Band G, Wilson JF, Capelli C, Falush D, Myers S. (2014). A genetic atlas of human admixture history. Science. 14;343(6172):747-51.

Hida A, Kitamura S, Katayose Y, Kato M, Ono H, Kadotani H, Uchiyama M, Ebisawa T, Inoue Y, Kamei Y, et al. (2014). Screening of clock gene polymorphisms demonstrates association of a PER3 polymorphism with morningness-eveningness preference and circadian rhythm sleep disorder. Sci Rep. 4:6309.

Hohjoh H, Takasu M, Shishikura K, Takahashi Y, Honda Y, Tokunaga K. (2003). Significant association of the arylalkylamine N-acetyltransferase (AA-NAT) gene with delayed sleep phase syndrome. Neurogenetics 4(3):151-3.

Honma S, Kawamoto T, Takagi Y, Fujimoto K, Sato F, Noshiro M, Kato Y, Honma K. (2002). Dec1 and Dec2 are regulators of the mammalian molecular clock. Nature. 419(6909):841-4.

Horne JA, Ostberg O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol.4(2):97-110.

Hu Y, Shmygelska A, Tran D, Eriksson N, Tung JY, Hinds DA. (2016). GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. Nat Commun. 2;7:10448.

Hut RA, Beersma DG. (2011). Evolution of time-keeping mechanisms: early emergence and adaptation to photoperiod. Philos Trans R Soc Lond B Biol Sci. 366(1574):2141-54.

Jagannath A, Hughes S, Abdelgany A, Pothecary CA, Di Pretoro S, Pires SS, Vachtsevanos A, Pilorz V, Brown LA, Hossbach M, et al. (2015). Isoforms of Melanopsin Mediate Different Behavioral Responses to Light. Curr Biol. 25(18):2430-4.

Jakobsson M, Rosenberg NA. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801-1806.

Jaramillo-Correa JP, Rodríguez-Quilón I, Grivet D, Lepoittevin C, Sebastiani F, Heuertz M, Garnier-Géré PH, Alía R, Plomion C, Vendramin GG, et al. (2015). Molecular proxies for climate maladaptation in a long-lived tree (Pinus pinaster Aiton, Pinaceae). Genetics 199(3):793-807.

Jeffreys H. (1961). Theory of probability. Clarendon Press.

Jones CR, Campbell SS, Zone SE, Cooper F, DeSano A, Murphy PJ, Jones B, Czajkowski L, Ptácek LJ. (1999). Familial advanced sleep-phase syndrome: A short-period circadian rhythm variant in humans. Nat Med. 5(9):1062-5.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 484(7392):55-61.

Jones KH, Ellis J, von Schantz M, Skene DJ, Dijk DJ, Archer SN. ( 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. 6(1):12-6.

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA. (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet. 66(3):979-88.

Kang JI, Park CI, Namkoong K, Kim SJ. (2015). Associations between polymorphisms in the NR1D1 gene encoding for nuclear receptor REV-ERBα and circadian typologies.Chronobiol Int. 32(4):568-72.

Katzenberg D, Young T, Finn L, Lin L, King DP, Takahashi JS Mignot E. (1998). A CLOCK polymorphism associated with human diurnal preference. Sleep 21(6):569-576.

Keller SR, Levsen N, Olson MS, Tiffin P. (2012). Local adaptation in the flowering-time gene network of balsam poplar, Populus balsamifera L. Mol Biol Evol. 29(10):3143-3152.

(book) Kimura, M. 1985. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Ko CH, Takahashi JS. (2006). Molecular components of the mammalian circadian clock. Hum Mol Genet. 5 Spec No 2:R271-7. Review.

Ko CH, Yamada YR, Welsh DK, Buhr ED, Liu AC, Zhang EE, Ralph MR, Kay SA, Forger DB, Takahashi JS. (2010). Emergence of noise-induced oscillations in the central circadian pacemaker. PLoS Biol. 8(10):e1000513.

Kornmann B, Schaad O, Bujard H, Takahashi JS, Schibler U. (2007). System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. PLoS Biol. 5(2):e34.

Kripke DF, Rex KM, Ancoli-Israel S, Nievergelt CM, Klimecki W, Kelsoe JR. (2008). Delayed sleep phase cases and controls. Journal of Circadian Rhythms.;6:6.

Kyriacou CP, Peixoto AA, Sandrelli F, Costa R, Tauber E. (2008). Clines in clock genes: fine-tuning circadian rhythms to the environment.Trends Genet. 24(3):124-32.

Kudielka BM, Federenko IS, Hellhammer DH, Wüst S. (2006). Morningness and eveningness: the free cortisol rise after awakening in "early birds" and "night owls". Biol Psychol. 72(2):141-6.

Iwase T, Kajimura N, Uchiyama M, Ebisawa T, Yoshimura K, Kamei Y, Shibui K, Kim K, Kudo Y, Katoh M, et al. (2002). Mutation screening of the human Clock gene in circadian rhythm sleep disorders. Psychiatry Res. 109(2):121-8.

Lamia KA, Storch KF, Weitz CJ. (2008). Physiological significance of a peripheral tissue circadian clock. Proc Natl Acad Sci U S A. 105(39):15172-7.

Lane JM, Vlasac I, Anderson SG, Kyle SD, Dixon WG, Bechtold DA, Gill S, Little MA, Luik A, Loudon A, et al. (2016). Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. Nat Commun. 7:10889.

Langmesser S, Tallone T, Bordon A, Rusconi S, Albrecht U. (2008). Interaction of circadian Clk proteins PER2 and CRY with BMAL1 and CLK. BMC Mol Biol. 9:41.

Lawson DJ, Hellenthal G, Myers S, Falush D (2012). Inference of Population Structure using Dense Haplotype Data. PLoS Genet 8(1): e1002453.

Lázár AS, Slak A, Lo JC, Santhi N, von Schantz M, Archer SN, Groeger JA, Dijk DJ. (2012). Sleep, diurnal preference, health, and psychological well-being: a prospective single-allelic-variation study. Chronobiol Int. 29(2):131-46.

Le Martelot G, Claudel T, Gatfield D, Schaad O, Kornmann B, Lo Sasso G, Moschetta A, Schibler U. (2009). REV-ERBalpha participates in circadian SREBP signaling and bile acid homeostasis. PLoS Biol. 7(9):e1000181.

Lee H, Chen R, Lee Y, Yoo S, Lee C. (2009). Essential roles of CKIdelta and CKIepsilon in the mammalian circadian clock. Proc Natl Acad Sci U S A. 106(50):21359-21364.

Lee HJ, Kim L, Kang SG, Yoon HK, Choi JE, Park YM, Kim SJ, Kripke DF. (2011). PER2 variation is associated with diurnal preference in a Korean young population. Behav Genet. 41(2):273-7.

Lee J, Lee Y, Lee MJ, Park E, Kang SH, Chung CH, Lee KH, Kim K. (2008). Dual modification of BMAL1 by SUMO2/3 and ubiquitin promotes circadian activation of the CLOCK/BMAL1 complex. Mol Cell Biol. 28(19):6056-65.

Li J, Li H, Jakobsson M, Li S, Sjödin P, Lascoux M. (2012). Joint analysis of demography and selection in population genetics: where do we stand and where could we go? Mol Ecol. 21(1):28-44.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science 319(5866):1100-1104.

Liu N, Zhao H. (2006). A non-parametric approach to population structure inference using multilocus genotypes. Hum Genomics. 2(6):353-64.

Long JC. (1991). The genetic structure of admixed populations. Genetics. 127(2):417-28.

Lowrey PL, Shimomura K, Antoch MP, Yamazaki S, Zemenides PD, Ralph MR, Menaker M, Takahashi JS. (2000). Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. Science. 288(5465):483-92.

Lowrey PL, Takahashi JS. (2004). Mammalian circadian biology: elucidating genome-wide levels of temporal organization. Annu Rev Genomics Hum Genet.5:407-41. Review.

Lowrey PL, Takahashi JS. (2011). Genetics of circadian rhythms in Mammalian model organisms. Adv Genet. 74:175-230. Review.

Lucas RJ. (2013). Mammalian inner retinal photoreception. Curr Biol. 23(3):R125-33.

Marcheva B, Ramsey KM, Buhr ED, Kobayashi Y, Su H, Ko CH, Ivanova G, Omura C, Mo S, Vitaterna MH, Lopez JP, Philipson LH, Bradfield CA, Crosby SD, JeBailey L, Wang X, Takahashi JS, Bass J. (2010). Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. Nature. 466(7306):627-31.

Marcheva B, Ramsey KM, Peek CB, Affinati A, Maury E, Bass J. (2013). Circadian clocks and metabolism. Handb Exp Pharmacol. (217):127-55. Review.

Mecacci L, Rocchetti G. (1998). Morning and evening types: stress-related personality aspects. Personality and Individual Differences 25:537-542.

Mohawk JA, Green CB, Takahashi JS. (2012). Central and Peripheral Circadian Clocks in Mammals. Annu Rev Neurosci. 35:445-462.

Mongrain, V, Cermakian, N. (2009). Clock genes in health and diseases. J Appl Biomed, 7, 15-33.

Mullen LM, Vignieri SN, Gore JA, Hoekstra HE. (2009). Adaptive basis of geographic variation: genetic, phenotypic and environmental differences among beach mouse populations. Proc Biol Sci. 276(1674):3809-18.

Nakahata Y, Kaluzova M, Grimaldi B, Sahar S, Hirayama J, Chen D, Guarente LP, Sassone-Corsi P. (2008). The NAD+-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control. Cell. 134(2):329-40.

Nandi D, Tahiliani P, Kumar A, Chandu D. (2006). The ubiquitin-proteasome system. J Biosci. 31(1):137-55. Review.

Numata H, Miyazaki Y, Ikeno T. (2015). Common features in diverse insect clocks. Zoological Lett. 1:10.

O'Malley KG, Banks MA. (2008). A latitudinal cline in the Chinook salmon (Oncorhynchus tshawytscha) Clock gene: evidence for selection on PolyQ length variants. Proc Biol Sci. 275(1653):2813-21.

O'Neill JS, Reddy AB. (2012). The essential role of cAMP/Ca2+ signaling in mammalian circadian timekeeping. Biochem Soc Trans. 40(1):44-50.

Ohsaki K, Oishi K, Kozono Y, Nakayama K, Nakayama KI, Ishida N. (2008). The role of ß-TrCP1 and ß-TrCP2 in circadian rhythm generation by mediating degradation of clock protein PER2. J Biochem. 144(5):609-18.

Okasha, Samir, "Population Genetics", The Stanford Encyclopedia of Philosophy (Fall 2015 Edition), Edward N. Zalta (ed.)

Ojeda DA, Perea CS, Niño CL, Gutiérrez RM, López-León S, Arboleda H, Camargo A, Adan A, Forero DA. (2013). A novel association of two non-synonymous polymorphisms in PER2 and PER3 genes with specific diurnal preference subscales. Neurosci Lett. 553:52-6.

Osland TM, Bjorvatn BR, Steen VM, Pallesen SI. (2011). Association study of a variable-number tandem repeat polymorphism in the clock gene PERIOD3 and chronotype in Norwegian university students. Chronobiol Int. 28(9):764-70.

Pandi-Perumal SR, Smits M, Spence W, Srinivasan V, Cardinali DP, Lowe AD, Kayumov L. (2007). Dim light melatonin onset (DLMO): a tool for the analysis of circadian phase in human sleep and chronobiological disorders. Prog Neuropsychopharmacol Biol Psychiatry. 31(1):1-11.

Parra EJ. (2007). Human pigmentation variation: evolution, genetic basis, and implications for public health. Am J Phys Anthropol. Suppl 45:85-105. Review.

Partch CL, Green CB, Takahashi JS. (2014). Molecular architecture of the mammalian circadian clock. Trends Cell Biol. 24(2):90-9.

Partonen T, Treutlein J, Alpman A, Frank J, Johansson C, Depner M, Aron L, Rietschel M, Wellek S, Soronen P, et al. (2007). Three circadian clock genes Per2, Arntl, and Npas2 contribute to winter depression. Ann Med. 39(3):229-38.

Patin E, Laval G, Barreiro LB, Salas A, Semino O, Santachiara-Benerecetti S, Kidd KK, Kidd JR, Van der Veen L, Hombert JM, et al. (2009). Inferring the Demographic History of African Farmers and Pygmy Hunter–Gatherers Using a Multilocus Resequencing Data Set. PLoS Genet. 5(4): e1000448.

Pedrazzoli M, Louzada FM, Pereira DS, Benedito-Silva AA, Lopez AR, Martynhak BJ, Korczak AL, Koike Bdel V, Barbosa AA, D'Almeida V, et al. (2007). Clock polymorphisms and circadian rhythms phenotypes in a sample of the Brazilian population. Chronobiol Int. 24(1):1-8.

Pedrazzoli M, Secolin R, Esteves LO, Pereira DS, Koike Bdel V, Louzada FM, Lopes-Cendes I, Tufik S. (2010). Interactions of polymorphisms in different clock genes associated with circadian phenotypes in humans. Genet Mol Biol. 33(4):627-632.

Pendergast JS, Friday RC, Yamazaki S. (2010). Distinct functions of Period2 and Period3 in the mouse circadian system revealed by in vitro analysis. PLoS One 5(1):e8552.

Pereira DS, Tufik S, Louzada FM, Benedito-Silva AA, Lopez AR, Lemos NA, Korczak AL, D'Almeida V, Pedrazzoli M. (2005). Association of the length polymorphism in the human Per3 gene with the delayed sleep-phase syndrome: does latitude have an influence upon it? Sleep. 28(1):29-32.

(book) Pittendrigh CS. Circadian rhythms and the circadian organization of living systems. Cold Spring Harbor Symposia on Quantitative Biology: Volume XXV. Biological Clocks. New York: Cold Spring Harbor Press, 1960. pp. 159–184.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet. 81(3):559-575.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW Jr, Orlando L, Metspalu E, et al. (2014). Upper Paleolithic Siberian genome reveals dual ancestry of Native Americans. Nature. 505(7481):87-91

Ralph MR, Menaker M. (1988). A mutation of the circadian system in golden hamsters. Science. 241(4870):1225-7.

Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci U S A. 102(44):15942-15947.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW Jr, Rasmussen S, Moltke I, Albrechtsen A, Doyle SM, et al., (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature. 13;506(7487):225-9.

Reick M, Garcia JA, Dudley C, McKnight SL. (2001). NPAS2: An Analog of Clock Operative in the Mammalian Forebrain. Science 293(5529):506-509.

Reid KJ, Chang AM, Dubocovich ML, Turek FW, Takahashi JS, Zee PC. (2001). Familial advanced sleep phase syndrome. Arch Neurol. 58(7):1089-94.

Rencher AC. (2002). Multivariate Analysis of Variance, in Methods of Multivariate Analysis, Second Edition, John Wiley & Sons, Inc., New York, NY, USA.

Roberts RD, Kyllonen PC. (1999). Morningness-eveningness and intelligence: early to bed, early to rise will likely make you anything but wise! Pers Individ Dif.27(6):1123-33.

Roecklein KA, Wong PM, Franzen PL, Hasler BP, Wood-Vasey WM, Nimgaonkar VL, Miller MA, Kepreos KM, Ferrell RE, Manuck SB. (2012). Melanopsin gene variations interact with season to predict sleep onset and chronotype. Chronobiol Int. 29(8):1036-47.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. (2002). Genetic structure of human populations. Science 298(5602):2381-2385.

Roenneberg T, Merrow M. (2003). The network of time: understanding the molecular circadian system. Curr Biol. Mar 4;13(5):R198-207. Review.

Roenneberg T, Wirz-Justice A, Merrow M. (2003a). Life between clocks: daily temporal patterns of human chronotypes. J Biol Rhythms. Feb;18(1):80-90.

Roenneberg T, Daan S, Merrow M. (2003b). The art of entrainment. J Biol Rhythms. Jun;18(3):183-94. Review.

Roenneberg T, Dragovic Z, Merrow M. (2005). Demasking biological oscillators: properties and principles of entrainment exemplified by the Neurospora circadian clock. Proc Natl Acad Sci U S A. 102(21):7742-7.

Roenneberg T, Merrow M. (2005). Circadian clocks - the fall and rise of physiology. Nat Rev Mol Cell Biol. 6(12):965-71. Review.

Roenneberg T, Allebrandt KV, Merrow M, Vetter C. Social jetlag and obesity. Curr Biol. 2012 May 22;22(10):939-43.

Rutter J, Reick M, Wu LC, McKnight SL. (2001). Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. Science. 293(5529):510-4.

Sack RL, Brandes RW, Kendall AR, Lewy AJ. (2000). Entrainment of free-running circadian rhythms by melatonin in blind people. N Engl J Med. 343(15):1070-7.

Saini C, Morf J, Stratmann M, Gos P, Schibler U. (2012). Simulated body temperature rhythms reveal the phase-shifting behavior and plasticity of mammalian circadian oscillators. Genes Dev. 26(6):567-80.

Saini C, Brown SA, Dibner C. (2015). Human peripheral clocks: applications for studying circadian phenotypes in physiology and pathophysiology. Front Neurol. 6:95.

Sahar S, Sassone-Corsi P. (2013). The epigenetic language of circadian clocks. Handb Exp Pharmacol. (217):29-44. Review.

Sato TK, Panda S, Miraglia LJ, Reyes TM, Rudic RD, McNamara P, Naik KA, FitzGerald GA, Kay SA, Hogenesch JB. (2004). A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. Neuron. 43(4):527-37.

Sawyer LA, Sandrelli F, Pasetto C, Peixoto AA, Rosato E, Costa R, Kyriacou CP. (2006). The period gene Thr-Gly polymorphism in Australian and African Drosophila melanogaster populations: implications for selection. Genetics 174(1):465-480.

Seeb LW, Waples RK, Limborg MT, Warheit KI, Pascal CE, Seeb JE. (2014). Parallel signatures of selection in temporally isolated lineages of pink salmon. Mol Ecol. 23(10):2473-2485.

Sehgal A, Mignot E. (2011). Genetics of sleep and sleep disorders. Cell. 146(2):194-207. Review.

So AY, Bernal TU, Pillsbury ML, Yamamoto KR, Feldman BJ. (2009). Glucocorticoid regulation of the circadian clock modulates glucose homeostasis. Proc Natl Acad Sci USA. 106(41):17582-7.

Stokkan KA, Yamazaki S, Tei H, Sakaki Y, Menaker M. (2001). Entrainment of the circadian clock in the liver by feeding. Science. 291(5503):490-3.

Sweeney BM, Hastings JW. (1960). Effects of temperature upon diurnal rhythms. Cold Spring Harb Symp Quant Biol. 25:87-104.

Takano A, Uchiyama M, Kajimura N, Mishima K, Inoue Y, Kamei Y, Kitajima T, Shibui K, Katoh M, Watanabe T, 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(10):1901-9.

Takeda N, Maemura K. (2015). The role of clock genes and circadian rhythm in the development of cardiovascular diseases. Cell Mol Life Sci. 72(17):3225-34.

Tauber E, Zordan M, Sandrelli F, Pegoraro M, Osterwalder N, Breda C, Daga A, Selmin A, Monger K, Benna C, et al. (2007). Natural Selection Favors a Newly Derived timeless Allele in Drosophila melanogaster. Science. 316(5833):1895-8.

Terman M, White TM. (2001). Morningness-Eveningness Questionnaire—Self-assessment version (AutoMEQSA). Retrieved from www.cet.org

Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J et al. (2001). Haplotype diversity and linkage disequilibrium at human G6PD:Recent origin of alleles that confer malarial resistance. Science 293(5529):455-462.

Toh KL, Jones CR, He Y, Eide EJ, Hinz WA, Virshup DM, Ptácek LJ, Fu YH. (2001). An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. Science 291(5506):1040-3.

Turek FW, Joshu C, Kohsaka A, Lin E, Ivanova G, McDearmon E, Laposky A, Losee-Olson S, Easton A, Jensen DR, et al. (2005). Obesity and metabolic syndrome in circadian Clock mutant mice. Science. 308(5724):1043-5.

Utge SJ, Soronen P, Loukola A, Kronholm E, Ollila HM, Pirkola S, Porkka-Heiskanen T, Partonen T, Paunio T. (2010). Systematic analysis of circadian genes in a population-based sample reveals association of TIMELESS with depression and sleep disturbance. PLoS One 5(2):e9259.

Viola AU, Archer SN, James LM, Groeger JA, Lo JC, Skene DJ, von Schantz M, Dijk DJ. (2007). PER3 polymorphism predicts sleep structure and waking performance. Curr Biol. 17(7):613-618.

Vitaterna MH, Takahashi JS, Turek FW. (2001). Overview of circadian rhythms. Alcohol Res Health. 25(2):85-93. Review.

Vitti JJ, Grossman SR, Sabeti PC. (2013). Detecting natural selection in genomic data. Annu Rev Genet.47:97-120.

Voinescu BI. (2009). Clock genes, chronotypes and diseases. HVM Bioflux 1(1):19-35.

von Schantz M. (2008). Phenotypic effects of genetic variability in human clock genes on circadian and sleep parameters. J Genet. 87(5):513-9.

Xu Y, Padiath QS, Shapiro RE, Jones CR, Wu SC, Saigoh N, Saigoh K, Ptácek LJ, Fu YH. (2005). Functional consequences of a CKIdelta mutation causing familial advanced sleep phase syndrome. Nature 434(7033):640-4.

Xu Y, Toh KL, Jones CR, Shin JY, Fu YH, Ptácek LJ. (2007). Modeling of a human circadian mutation yields insights into clock regulation by PER2. Cell.128(1):59-70.

Yamazaki S, Numano R, Abe M, Hida A, Takahashi R, Ueda M, Block GD, Sakaki Y, Menaker M, Tei H. (2000). Resetting central and peripheral circadian oscillators in transgenic rats. Science. 288(5466):682-5.

Yoo SH, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, Buhr ED, Siepka SM, Hong HK, Oh WJ, Yoo OJ, et al. (2004). PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc Natl Acad Sci U S A.101(15):5339-46.

Young MW, Kay SA. (2001). Time zones: a comparative genetics of circadian clocks. Nat Rev Genet. 2(9):702-15. Review.

Yule GU. (1902). Mendel's laws and their probable relations to intra-racial heredity. New Phytol. 1(193-207):222–238.

Wallach T, Kramer A. (2015). Chemical chronobiology: Toward drugs manipulating time. FEBS Lett. 589(14):1530-8.

Welsh DK, Takahashi JS, Kay SA. (2010). Suprachiasmatic nucleus: cell autonomy and network properties. Annu Rev Physiol.;72:551-77.

Weinberg W. (1908). Über den Nachweis der Vererbung beim Menschen. Jahresh. Ver. Vaterl. Naturkd. Württemb. 64 369–382 (English translations in Boyer 1963 and Jameson 1977)

Wilkinson KA, Henley JM. (2010). Mechanisms, regulation and consequences of protein SUMOylation. Biochem J. 428(2):133-45.

(tutorial) Zheng X. (2013). A Tutorial for the R Package SNPRelate. GENEVA Coordinating Center Department of Biostatistics University of Washington

Zhou M, Kim JK, Eng GW, Forger DB, Virshup DM. (2015). A Period2 Phosphoswitch Regulates and Temperature Compensates Circadian Period. Mol Cell. 60(1):77-88.

Zhu Y, Stevens RG, Leaderer D, Hoffman A, Holford T, Zhang Y, Brown HN, Zheng T. (2008). Non-synonymous polymorphisms in the circadian gene NPAS2 and breast cancer risk. Breast Cancer Res Treat. 107(3):421-5.

Zhu L, Zee PC. (2012). Circadian rhythm sleep disorders. Neurol Clin. 30(4):1167-91.

Zisapel N. (2001). Circadian rhythm sleep disorders: pathophysiology and potential approaches to management. CNS Drugs. 15(4):311-28. Review.

Zylka MJ, Shearman LP, Weaver DR, Reppert SM. (1998). Three period homologs in mammals: differential light responses in the suprachiasmatic circadian clock and oscillating transcripts outside of brain. Neuron. 20(6):1103-10.