Molecular Characterization of a Diurnal Rodent *Funambulus Palmarum* (South Indian Palm Squirrel): Cloning and characterization of *Period 2* from the Suprachiasmatic Nucleus

A thesis submitted to the University Of Hyderabad for the award of

Doctor of Philosophy

in

Animal Sciences

By

Varsha S. Prasad

Department of Animal Biology

School of Life Sciences

University Of Hyderabad

Hyderabad- 500 046.

India

Enrollment No: 10LAPH07

August 2016

UNIVERSITY OF HYDERABAD

(A Central University Established in 1974 By An Act Of Parliament)

HYDERABAD - 500 046, INDIA

DECLARATION

I, **Varsha S. Prasad**, hereby declare that this thesis entitled "**Molecular Characterization of Diurnal Rodent *Funambulus palmarum* (South Indian Palm Squirrel): Cloning and characterization of *Period 2* from the Suprachiasmatic nucleus**" submitted by me under the guidance and supervision of Dr. Anita Jagota, is an original and independent research work. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma.

Date: Name: Varsha S. Prasad

Signature:

Reg. No: 10LAPH07

UNIVERSITY OF HYDERABAD
(A Central University established in 1974 By The Act Of Parliament)
HYDERABAD – 500 046, INDIA

CERTIFICATE

This is to certify that this thesis entitled **"Molecular Characterization of Diurnal Rodent *Funambulus palmarum* (South Indian Palm Squirrel): Cloning and characterization of *Period 2* from the Suprachiasmatic nucleus"** is a record of bonafide work done by Ms. Varsha S. Prasad, a research scholar for Ph.D. programme in Animal Sciences, School of Life Sciences, University of Hyderabad under my guidance and supervision.

The thesis has not been submitted previously in part or in full to this or any other University or Institution for the award of any degree or diploma.

Dr. Anita Jagota

Supervisor
Associate Professor
Dept. of Animal Biology

Head

Dean

Department of Animal Biology School of Life Sciences

Acknowledgements

I express my profound sense of gratitude to my supervisor **Dr. Anita Jagota** for her relentless encouragement, guidance and support. She as an individual has always appreciated and respected my individuality for which I am very thankful.

I am grateful to Head, Dept. of Animal Biology **Prof. Jagan Pongubala** and former Heads, **Prof. B. SenthilKumaran**, **Prof. Manjula Sritharan** and **Prof. S. Dayananda** for allowing me to use the Department facilities.

I am thankful to Dean, School of Life Sciences, **Prof. P. Reddanna** and former Deans **Prof. R.P. Sharma**, **Prof. Aparna Dutta Gupta**, **Prof. M. Ramanadham** and **Prof. A.S. Raghavendra** for allowing me to use school facilities.

I am grateful to my Doctoral Committee members, **Prof. S. Dayananda** and **Prof. Anand K. Kondapi** for their critical assessment, discussions and advise during my research work.

I thank faculty members of School of Life Sciences for their timely and useful suggestions.

I thank all my lab mates and good friends **Dr. Pavan**, **Dr. Kapil**, **Dr. Yalamanda**, **Dr. Anumodh**, **Dr. Ushodaya**, **Dr. Dileep**, **Dr. Lakshmi**, **Dr. Vinod**, **Kunal**, **Neelesh**, **Kowshik**, **Pratik**, **Dhivya**, **Minurani** and **Pratibha** for their timely help and in keeping a very pleasant atmosphere in the laboratory.

I thank **Mr. Gopal Krishna Pradhan**, **Mr. Mallesh**, **Mr. Satish** and **Mr. Balakrishna** for technical help in the lab.

Mr. Balakrishna is specially thanked for his timely and skilful involvement in maintaining the animal house and laboratory.

I would like to thank all Department staff **Mr. Ankinedu**, **Ms. Vijaya Lakshmi**, **Ms Jyoti**, **Mr. Jagan**, **Mr. Babu Rao, Mr. Gopi** and **Mr. Srinivas** for their help in office work.

I thank staff of Animal House for their help whenever was needed.

I would also like to thank all the Research Scholars in School of Life Sciences for their cooperation and help.

I duly acknowledge the School's Engineering and Electrical division staffs for being extremely supportive.

I thank **DST**, **CSIR**, **ILS**, **UPE-I** and **II**, **CREA**, **DST**, **DST-Nano**, **ICMR**, **UGC**, **CWC** and **DBT** for funding to the laboratory.

I thank **DBT**, **DBT-CREBB**, **CSIR**, **DST**, **INSA**, **DST-FIST**, **UGC-SAP** for funding to Department and School.

I specially thank **UoH BBL** for fellowship.

I sincerely thank all my teachers, past and present.

I am very lucky to have a very selflessly supportive family who has always respected my decisions and loved me unconditionally.

My husband has shown equal interest in my career, helping and supporting me at every phase of my work, without which it would have been a difficult journey. I also thank my Parents-by-Law for being very understanding and respecting the demands of a student.

I thank my 'Achan' and 'Amma' for everything that I am today and will be tomorrow.

CONTENTS

Introduction and Review of Literature

Circadian rhythms and cellular oscillators

Biological clocks are time keeping mechanisms that evolved in organisms allowing them to adapt and anticipate the various geophysical variations in their external habitat. The day-night cycle caused by Earth's rotation on its axis has been the primary selective pressure to impose a rhythmic pattern to physiological and behavioral activities that nearly reflect the length of a day on earth. Such oscillations with a near (circa) 24h periodicity termed as circadian rhythms allows species- specific partitioning of specific activities across different parts of a diel (day) cycle conferring survival advantage to the bearer. Though light act as the major 'Zeitgeber' or 'time giver' to synchronize or 'entrain' the phase of these rhythms to that of the zeitgeber time (ZT), they also persists independent of any photic variable and therefore constitutes the hands of the clock which dictates the 'circadian or endogenous time' (CT) of any living entity. The internal 'period' varies across individuals of the same species but it is relatively consistent over a range of temperatures (temperature compensation). Apart from circadian rhythms cycles having shorter or longer intrinsic periods are also biologically significant. Ultradian rhythms, have periodicity of less than 20h whereas infradian rhythms are longer than 28h, they can be measured in weeks, months, years (circannual), and longer.

The central organizing principle of the clockwork has been the presence of a cellular autonomous circadian oscillator despite of the varying phylogenetic histories and biological complexities among the organisms reported to have a circadian clock, Basically, every cell is an oscillator that comprises of 'positive' and 'negative' elements that encodes within its interactions the ticking of an internal clock. The positive elements activate the synthesis of negative elements which in turn negatively feeds back to inactivate its own production; with activation and inhibition separated by a time delay that constitutes the circa 24h endogenous period of the oscillator (Fig.1). Interlocking positive and negative loops between the clock components confer robustness and stability to the oscillator (Pederson et al., 2005). The table given below shows the components of the molecular clock in eukaryotic species.

Fig.1.

Basic clock components in eukaryotic species.

Species	Positive elements	Negative elements	Other known clock components
Neurospora	WC-1; WC-2	FRQ	VVD, FRH
Arabidopsis	TOC1; LUX; ELF4	CCA1/LHY	GI; PRR3,5,7,9; ELF3
Drosophila	CYC; CLK	PER; TIM	PDP1ε; VRI; CWO
Vertebrates	BMAL1; CLOCK	PER1,2; CRY1,2	RORα, β; REV-ERBα, β; DEC1, 2; DBP; E4BP4

Zang et al., 2011

Another unifying feature among the clock elements across taxa is the presence of a PAS module in their structure. PAS is an acronym formed from the names of the proteins in which these domains were first recognized, the Drosophila period clock protein (PER), vertebrate aryl hydrocarbon receptor nuclear translocator (ARNT), and Drosophila single-minded protein (SIM). They are sensory structures adapted to monitor changes in light, redox potentials, oxygen and energy levels within cells (Taylor and Zulin,1999). In vertebrates the interaction between the 'positive' and 'negative' elements occurs through a Transcriptional –Translational Feed (TTFL) back loop where subjective day is marked by transcriptional activation of the 'negative elements' Period (*Per* 1, 2 and 3) and Cryptochrome (Cry 1 and 2) genes by bHLH-PAS transcription factor (TFs) CLOCK and BMAL1 heterodimer binding to their respective 'E' box promoter elements (Fig.2).

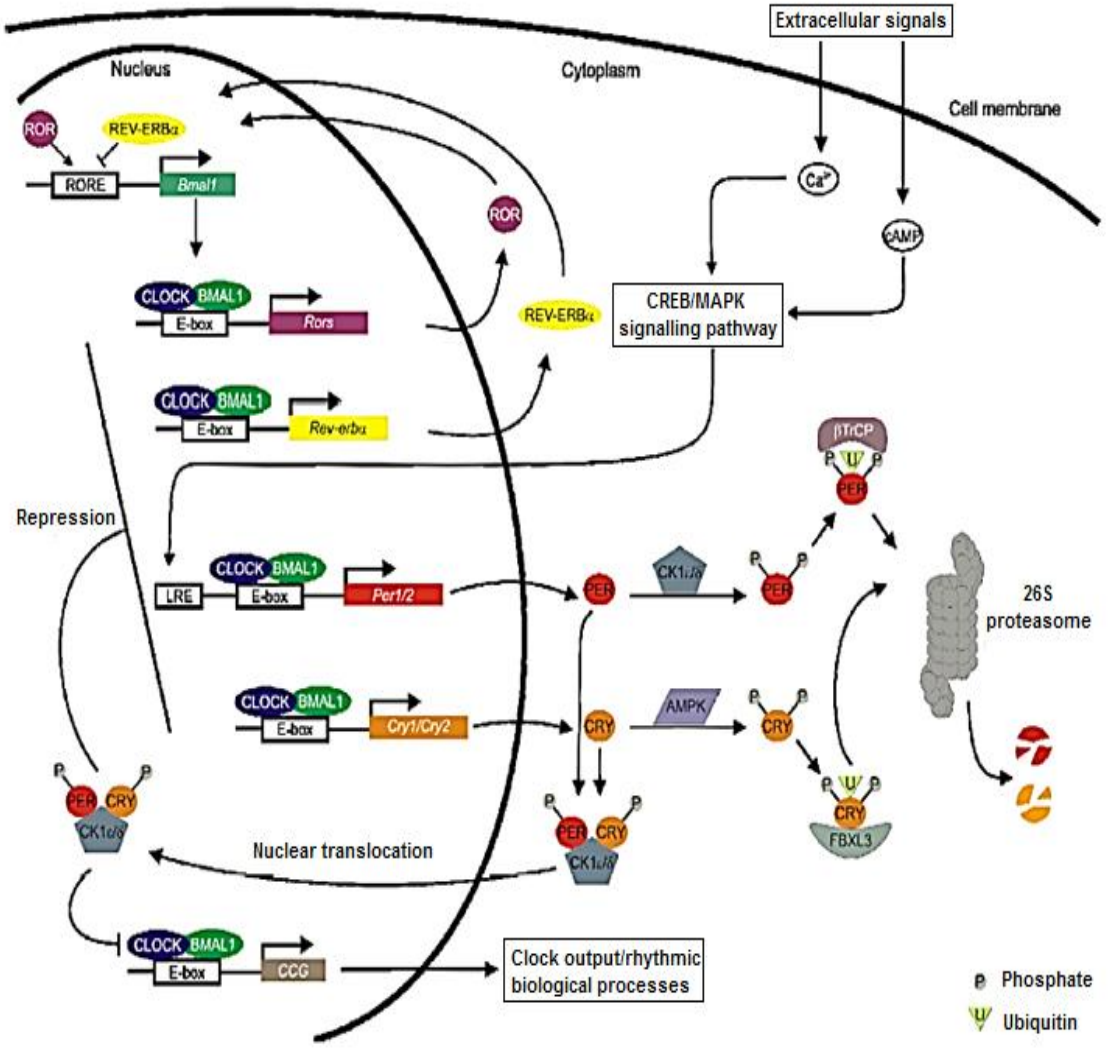


Fig.2. Model of the mammalian cell-autonomous oscillator. CCG, clock-controlled gene; P, phosphate; U, ubiquitin.

Lowrey and Takahashi, 2011

PER and CRY protein complex accumulates in the cytoplasm, and at a critical concentration during the subjective dawn translocates to the nucleus where they bind to and repress CLOCK/BMAL1 activity. The time delay programmed between transcriptional activation and repression by means of post-translational modifications regulates repressor turn over rate by dictating time specific protein-protein interactions, subcellular localization and proteasomal degradation (Ripperger and Albrecht, 2012). The core loop is further augmented and stabilized by accessory pathways involving orphan nuclear receptor proteins, REV-ERBα and ROR*α* through their differential impact on *Bmal1* transcription. The TF heterodimer also induces *Rev-erbα* transcription, the protein products so formed acts as a transcriptional repressor which competes with the transcriptional inducer, ROR for the *Bmal1* promoter (Ripperger and Brown, 2010). External stimuli or' Zeitgeber'communicates to the clock by way of 'inducible gene expression' and further down stream signalling events helps in integrating this time information to the oscillator time and also in synchronizing the output of multiple oscillators enabling stable entrainment.

The Suprachiasmatic nucleus (SCN) is the circadian pacemaker in mammals

The collective outputs of cellular oscillators are functionally integrated and co-ordinated at the organ and organismal level by a central oscillator or the 'Master clock'. In mammals the hypothalamic Suprachiasmatic nucleus (SCN) was proved to be the master circadian pacemaker after decades of research involving surgical and electrolytic ablations of these neurons resulting in total abolishment of physiological rhythms, which could be reestablished by allotransplantion. SCN cultures can remain rhythmic for over an year unlike peripheral clocks ascertaining the autonomous nature of rhythm generation (Silver and Rainbow, 2013). The Afferent pathways which carry photic as well as non-photic 'cues' to the SCN and the Efferent pathways that conveys temporal information to synchronize 'peripheral' clock rhythms constitute the 'tripartite organization' of the Circadian timing System (Fig.3). Melanopsin expressing retinal ganglionic cells in the Retinohypothalamic tract (RHT) which are more sensitive to the 'non image forming' region of the photic spectra directly conveys light information to the SCN. Glutamate and Pituitary adenylate cyclase-activating peptide (PACAP) are the major neurotransmitters of this pathway. The other neuronal tracts carrying photic information reach the SCN indirectly, they are Geniculohypothalamic (GHT) and the Retino Raphe pathway. The Retino raphe pathway mainly carry non photic cues and have Serotonin as the major neurotransmitter, the GHT is rich in Neuropeptide Y (NPY) (Jagota, 2006).

The SCN in a sagittal view is located in the anterior hypothalamus. In the coronal plane, it is made up of a pair of nuclei located in the anteroventral hypothalamus, above the optic chiasm lying on either side of the third ventricle (Silver and Rainbow, 2013). SCN contains roughly 10,000 neurons each on either side, this number can vary across different organisms (Fig.4) Neurochemical and functional studies indicates that SCN is

a complex assemblage of multiple oscillators that are coupled in-phase by means of electrical and neurochemical signaling.

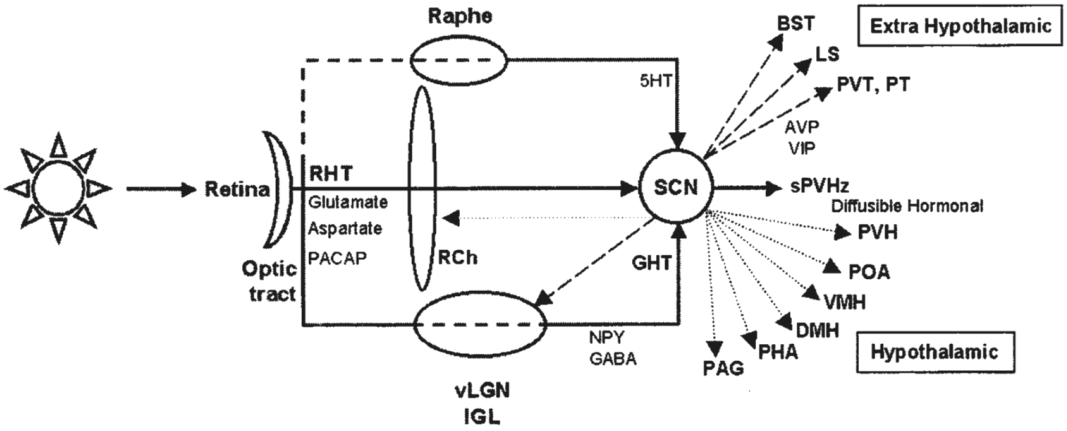
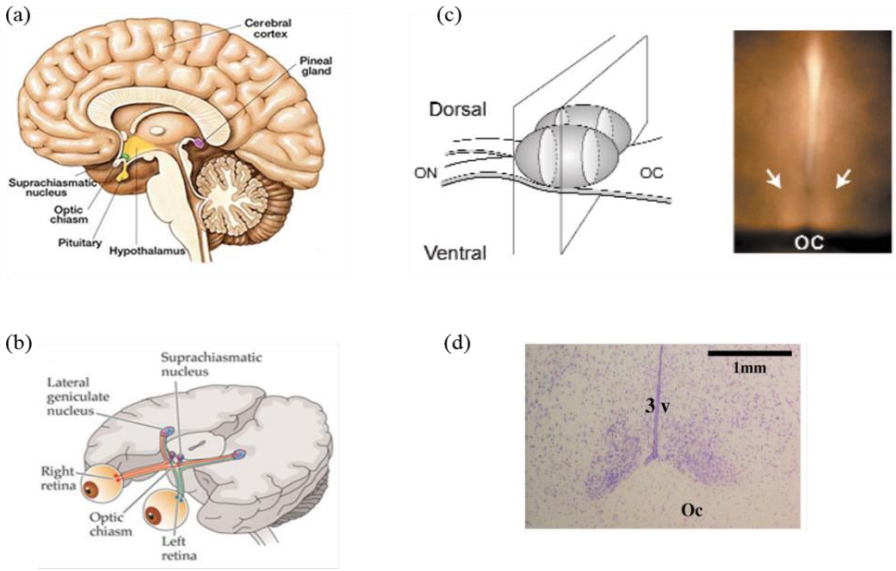


Fig.3. Diagrammatic representation of the (1) Photic afferents of SCN: RHT-retinohypothalamic tract; GHT-geniculohypothalamic tract via ventro lateral geniculate nucleus (vLGN) and the intergeniculate nucleus (IGL); from median dorsal raphe (ii) Efferents of the SCN towards various extrahypothalamic and hypothalamic targets: BST, bed nuclei of the striata terminalis; LS, lateral septal nucleus; PVT/PT, paraventricular nucleus of the thalamus, paratanial nucleus; IGL, intergeniculate leaflet; AHA, anterior hypothalamic area; PVH, paraventricular nucleus of hypothalamus; MPOA and POA, preoptic area nuclei; RCh, retrochiasmatic area; VMH, ventromedial nucleus of the hypothalamus; DMH, dorsomedial nucleus of hypothalamus; ZI, zonz incerta; PHA, posterior hypothalamic area; PAG, periaqueductal gray.

Jagota, 2006



https://ankiweb.net; Jagota et al., 2000; Mammen and Jagota, 2011

Fig. 4 (a) & (b) Diagrammatic representation of Brain sections showing the position of SCN. (c) SCN as observed through a dissection microscope in coronal sections (orientation depicted in the left diagram), OC-optic chiasm. (d) SCN in NeuN stained coronal section, 3v-3rd ventricle.

The multi oscillatory concept of the SCN

The multi oscillatory nature of the neurons was made evident by experiments involving uncoupling agents such as constant light (de la Iglesia et al., 2000) or surgical interventions that could alter neuronal signaling pattern, such as the plane of tissue sectioning in Jagota et al., (2000). Splitting of activity rhythms in hamsters under LL conditions as well as the two distinct bouts of electrical activity in SCN horizontal sections reflects the uncoupling of dual oscillators which free runs with independent periodicities in presence of desynchronizing agents (Fig.5). The heterogeneity is further established on the basis of RHT innervation and distribution of specific neurotransmitters. At this level SCN can be distinguished into 2 regions, namely the retino recipient, Vasoactive Neuropeptide (*Vip*) enriched 'core' which is ventrolateral and the dorso medial non retino-recipient 'shell' marked by neurons expressing Arginine vasopressin (AVP) (Fig.6). VIP is one of the most important coupling agents responsible for synchrony in the SCN. In mice, the abolition of VIP-based signaling leads to weak rhythms or arrythmicity in behavior, clock protein oscillations, and firing rate in SCN explants (Ananthasubramaniam et al., 2014). The shell neurons generate high amplitude oscillations of clock gene expression which is responsible for the endogenous clock period. Whereas the core neurons are thought to act as an integrator of external input that has to be communicated to the rest of the SCN. A relatively low amplitude clock gene rhythms in these neurons is hence an adaptation to sense environmental perturbations (Colwell, 2011; Zhu et al., 2012).

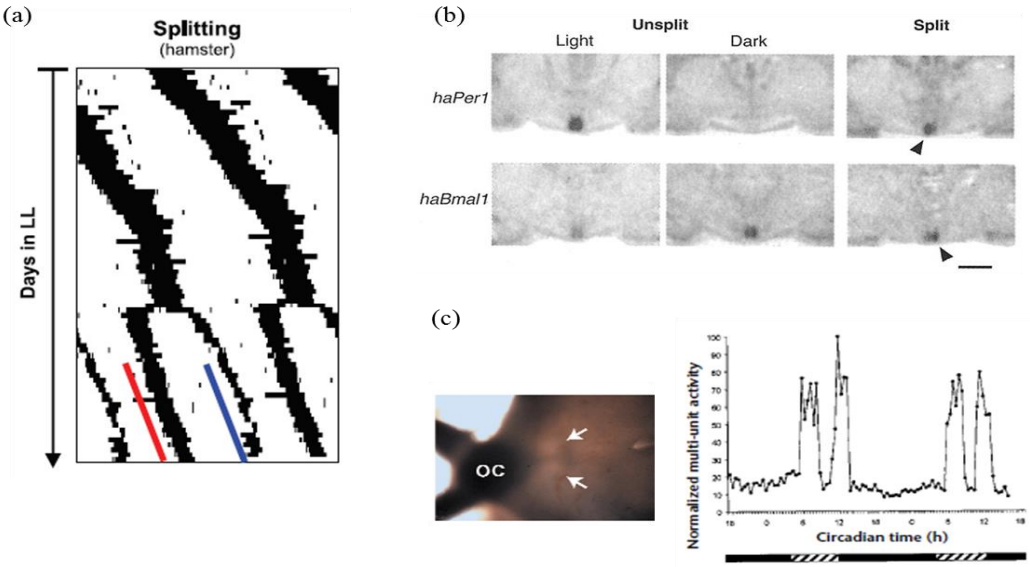


Fig.5. Experiments which demonstrated the multi-oscillatory concept of SCN

(a) Diagrammatic representation of splitting of activity rhythms in hamster under LL (b) mRNA characteristic of subjective day (*Per1*) and subjective night (*Bmal1*) simultaneously expressed on opposite sides of paired SCN in behaviorally split hamsters indicating uncoupling of oscillators (c) Rhythms of SCN electrical activity recorded for 48h. The two bouts of electrical firing correspond to uncoupled M and E oscillators, respectively.

Golombek and Rosenstein, 2010; de la Iglesia, 2000; Jagota et al., 2000

Photic response in SCN is temporally gated

The response of SCN towards light is biphasic, photic stimuli during the early subjective night causes phase advance while during the late subjective night it results in phase delay of behavioral rhythms (Fig.7). SCN is unresponsive to photic stimuli during subjective day. The pacemaker neurons are in a highly depolarized state during the subjective day, light which mediates its effect via Ca^{2+} influx and membrane depolarization is unable to further alter the membrane potential at this time and hence exerts its effect when the membrane is hyperpolarized at subjective night (O'Neill and Reddy, 2012). Photic stimuli are received by the 'core' neurons and through Glutamate signaling impinges on CRE (cAMP Response Element) activation of clock genes. The downstream secondary signaling events may further help in communicating the 'phase change' to the rest of the SCN.

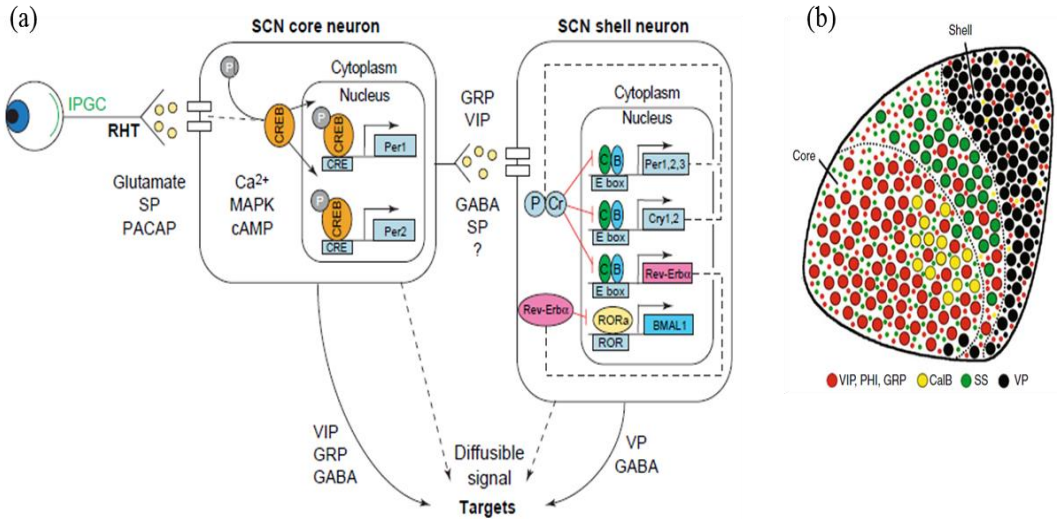


Fig.6. Rhythmic and Nonrhythmic compartments of SCN

(a) Light is transduced into a neural signal by intrinsically photoreceptive ganglion cells (IPGCs) in the retina and conveyed to the SCN core along the retinohypothalamic tract (RHT), resulting in the release of the neurotransmitter glutamate and the neuromodulators substance P (SP) and pituitary adenylyl cyclase activating peptide (PACAP) onto retino-recipient cells in the SCN core. Glutamate activates NMDA receptors, causing an influx of Ca^{2+}, which activates kinases such as mitogen-activated protein kinase (MAPK), resulting in phosphorylation of cAMP-response-element-binding protein (CREB). Activated CREB binds to the Ca^{2+}/cAMP response element (CRE) in the promoter region of both Per1 and Per2, activating their transcription. Neurons in the SCN core communicate with the rhythmic SCN shell and SCN targets using a variety of neurotransmitters, including vasoactive intestinal polypeptide (VIP), gastrin-releasing peptide (GRP) and SP. Additionally, almost all SCN cells are GABAergic. Cells in the rhythmic SCN shell contain molecular clocks driven by an autoregulatory transcription–translation loop. CLOCK (C) and BMAL (B) dimerize and bind to E-boxes in the promoter region of Period (Per) genes, Cryptochrome (Cry) genes and Rev-Erb*a*, activating their transcription. SCN shell neurons communicate with SCN targets using VP and GABA as neurotransmitters. Additionally, the SCN communicates with some target sites using a diffusible signal. (b) Neuropeptides of the SCN. The SCN of the hamster can be divided into a ventrolateral core and a dorsomedial shell, both densely packed with somata of small neurons. The main neuropeptide transmitters of these areas are vasoactive intestinal polypeptide (VIP), peptide histidine-isoleucin (PHI), gastrin-releasing peptide (GRP), calbindin (CalB), somatostatin (SS), and vasopressin (VP). Most neurons contain additionally GABA. The somata of the neurons are shown as larger circles, their projections as small circles.

Antle and Silver, 2005; Silver and Rainbow, 2013

Temporal harmony is indispensable for health and wellbeing.

Both neuronal and humoral signals originating from the SCN regulate the rhythmic output from peripheral organs that are 'slave' oscillators to the master clock and therefore can regulate various physiological functions such as timing of hormone release, feeding behavior, body temperature fluctuations etc. SCN also controls the rhythmic production and release of Melatonin from the pineal, thereby contributing to the circadian component of sleep propensity in humans (Takahashi et al., 2008). Given the indispensable role of SCN in orchestrating an adaptive internal temporal harmony, dysfunction of the clock can have serious impact on the health of its bearer. Apart from the classic circadian disturbances caused by desynchrony, such as jet lag syndrome, affective disorders, sleep disturbances etc. (Zee et al., 2013) the circadian component in many neurodegenerative disorders (Wulff et al., 2010; Reddy and Jagota, 2014; Mattam and Jagota, 2015) metabolic diseases(Maury et al., 2010; Marcheva et al., 2013) and addictive behaviors (Spanagel et al., 2005; Jagota and Reddy, 2007) is widely being acknowledged. Though not all medical disorders are a direct effect of a misaligned clock, the severity of their symptoms shows a circadian pattern. This knowledge is now utilized in planning treatment schedules which could have relatively better outcome when administered at specific time of a day. The idea of 'chronomedicine' which would offer personalized treatments tailored to individual circadian parameters can significantly augment the effectiveness of current therapeutic practices (Preubner and Heyd, 2016). Novel small-molecules that can modulate specific elements of the oscillator such as, clock proteins, kinases or epigenetic regulators can offer potential protective effects against several clock related disorders (He and Chen, 2016)

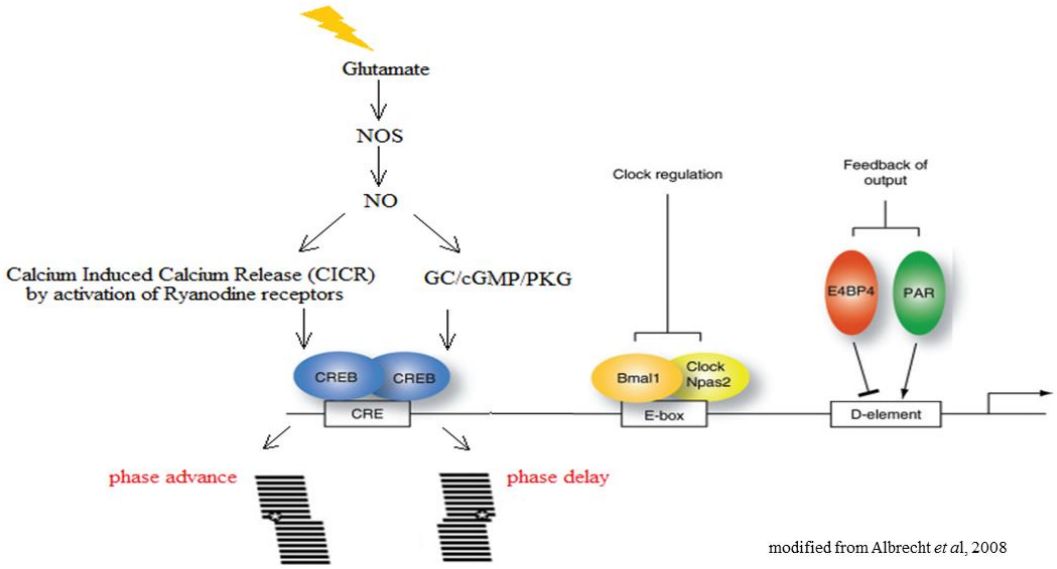


Fig. 7. Diagrammatic representation of a photo inducible clock gene with its promoter elements, CRE, E-box and D element. Light activates signaling cascade downstream of Glutamate induction. Based on the temporal profile of the photic cue the signaling bifurcates at the level of Nitric Oxide (NO) production, both the phase advancing and delaying light pulse activate CRE mediated transcript induction.

Challenges to the transcription-centric TTLF model for rhythm generation

Post-transcriptional mechanisms

The syntheses of clock proteins are delayed by several hours relative to their corresponding mRNAs. Post transcriptional mechanisms could therefore be crucial in regulating the sustenance of these transcripts as well as the timing of protein occurrence (Zang et al., 2011) The idea of the TTFL model as the core element of circadian clock has been challenged by recent Next Generation Sequencing (NGS) and Nascent seq data on circadian pre-mRNA expression where *de novo* transcription accounted for the rhythmicity of only 22% of transcripts. 80% of the clock controlled genes (ccg) mRNAs did not show rhythmic *de novo* transcription, while 50-70% that was rhythmic in this aspect was not rhythmic with respect to mRNA expression (Kojima and Green, 2015) Proteome analysis also shows that almost 50% of the rhythmically expressed proteins do not express rhythmicity in their mRNA levels (Mauvoisin et al., 2014). This lack of concordance between the mRNA and protein rhythms signifies the role of Posttranscriptional regulation in the functioning of molecular clock.

Regulation at different levels of pre-mRNA processing

Regulatory mechanism involved in different stages of pre-mRNA processing, such as capping, splicing, polyadenylation and cytoplasmic export as well as mRNA stability can contribute to the temporal profile of transcript and/or protein expression (Preubner and Heyd, 2016) (Fig.8). Circadian rhythm in RNA Polymerase II (RNA PolII) recruitment and initiation probably as a consequence of rhythmic transcription factor binding on promoter elements was reported by Koike et al., 2012. The presence of the 5' 7 methyl-guanosine 'cap' is essential for cap-dependent mRNA translation, splicing, polyadenylation mRNA stability and export in eukaryotes (Cowling, 2010). The enzyme that catalyzes 'cap methylation' namely RNA guanine-7 methyltransferase (RNMT) is recently known to have an indirect regulatory effect on circadian cycle. In cell lines, knock down of RNMT causes elongation of the circadian period by delaying the exit of mature mRNA from the nucleus proving the significance of RNA-methylation dependent RNA processing in setting the pace of the clock (Fustin et al., 2013). Using mouse exon-arrays McGlincy et al., 2012 identified exons that are alternatively spliced in a circadian manner in the liver. The circadian regulation of 62 RNA binding proteins (RBP) that may be involved in rhythmic processing of RNA was also revealed in the study. The mechanism of 'entrainment' involving 'transients' is now linked to the presence of a circadian and light inducible splicing switch producing a variant of the splicing factor U2AF26 mRNA. The alternate isoform of this protein that bears a C-terminal resemblance to the drosophila TIM binds to PER1 protein and attenuates its expression in mice. Limiting the expression of inducible genes is a buffering mechanism allowing slow and synchronized entrainment by preventing sudden and large changes to the clock (Jagannath et al., 2013). In addition, alternative splicing can regulate mRNA levels through the targeted degradation of isoforms having Premature Termination Codons (PTC) by nonsense-mediated decay (NMD). This

‘regulated unproductive splicing and translation (RUST)’ which can temporally regulate the abundance of full length transcripts are reported in *Arabidopsis* clock (Kwon et al., 2014).

Even when the steady-state mRNA levels are not rhythmic the circadian system can generate rhythmic protein synthesis also through rhythmic polyadenylation and deadenylation of the mRNA poly(A) tail. Regulation at the level of poly (A) length affects the translatability and stability of the transcript. ‘Poly (A) adenylome analysis’ shows that approximately 2.5% of mRNAs shows circadian fluctuation in the length of their poly(A) tail which correlated with the rhythmicity of the protein levels. (Kojima and Green, 2015). The circadian expression of the deadenylase Nocturnin (*Noc*, also called *Ccrn4l* [carbon catabolite repression 4-like]) with peak transcript levels at night has been reported in many tissues in mouse (Kojima et al., 2010). N6-methyladenosine (m6A) is the most common posttranscriptional modification at the Pu[G>A]m6AC[U>A>C] conserved motifs of an RNA, preferentially enriched near the stop codon, 3’UTR’s and within long internal exons. The dynamic nature of this RNA methylation is dictated by methyltransferases (METTL3-METTL14-WTAP complex) and demethylases (FTO) regulating its abundance, alternate splicing and translation by way of Ribonucleoproteins that ‘read’ these signals and trigger specific downstream activities. (Yue et al., 2015). Silencing *Mettl3* depletes m6A prolonging nuclear retention of *Per2* and *Arntl* resulting in lengthening of circadian period (Fustin et al., 2013), similar to the role of RNMT described earlier. Furthermore, the stability of mRNA is reported to be under circadian control. The mRNAs of the cycling transcripts *Per2, Per3* and *Cry1* are more stable during the rising phase of the rhythm in comparison to their declining phase (Kojima and Green, 2015). The rhythmicity in mRNA decay kinetics is attributed to circadian microRNAs (miRNAs), long non coding RNAs (LncRNAs) as well as the RNA binding Proteins that binds to the 3’UTR of clock transcripts differentially modulating alternative poyadenylation, poly(A) tail length, translational availability etc. miR-132 and miR-219 are directly involved in regulating circadian period and photic response (Cheng et al., 2007) Recent studies on Heterogenous Nuclear Ribonucleoproteins (hnRNP) such as hnRNP K and hnRNP D and Polypyrimidine tract Binding protein (PTB) implicates towards a role in circadian oscillation of *mPer3* (Kim et al., 2015).

Post transcriptional modifications aimed at promoting rhythmic translation also involves the rhythmic phosphorylation of the cap binding protein eIF4e in mammals controlling the rhythmic translation of *Per1* and *Per2* mRNA. The mechanism is integrated to photic induction via CREB dependent kinase activation involving MAPK and TORC1 pathways that phosphorylates and inhibits eIF-4E repressor protein 4E-BP1 (Cao et al., 2015). Cap dependent translation is also potentiated by rhythmic phosphorylation of BMAL1 which forms a complex with the eukaryotic Initiation factors in a circadian manner (Lipton et al., 2015). Moreover there is emerging significance of translational regulation mediated via different *cis*-acting elements in the RNA 5’ leader sequence of ccg transcripts in the regulation of temporal oscillation.

These elements include RNA secondary structure that inhibit ribosome scanning at the initial AUG, internal ribosome entry sites (IRES) mediating cap independent translation and upstream open reading frames (uORFs) which have an initiation codon in frame with a termination codon upstream or downstream of the main AUG (Lee et al., 2012; Barbosa et al., 2013; Janich et al., 2015). Rhythmic cap independent translation mediated by HNRNP-Q on the IRES of *mPer1* 5'UTR is thought to fine-tune the expression and amplitude of mPER1 independent of mRNA oscillation (Lee et al., 2011). IRES mediated translation is also reported to be important for the nocturnal production of Aryl alkylamine N-acetyltransferase (AANAT) protein, the rate limiting factor in melatonin synthesis from the pineal gland (Kim et al., 2007).

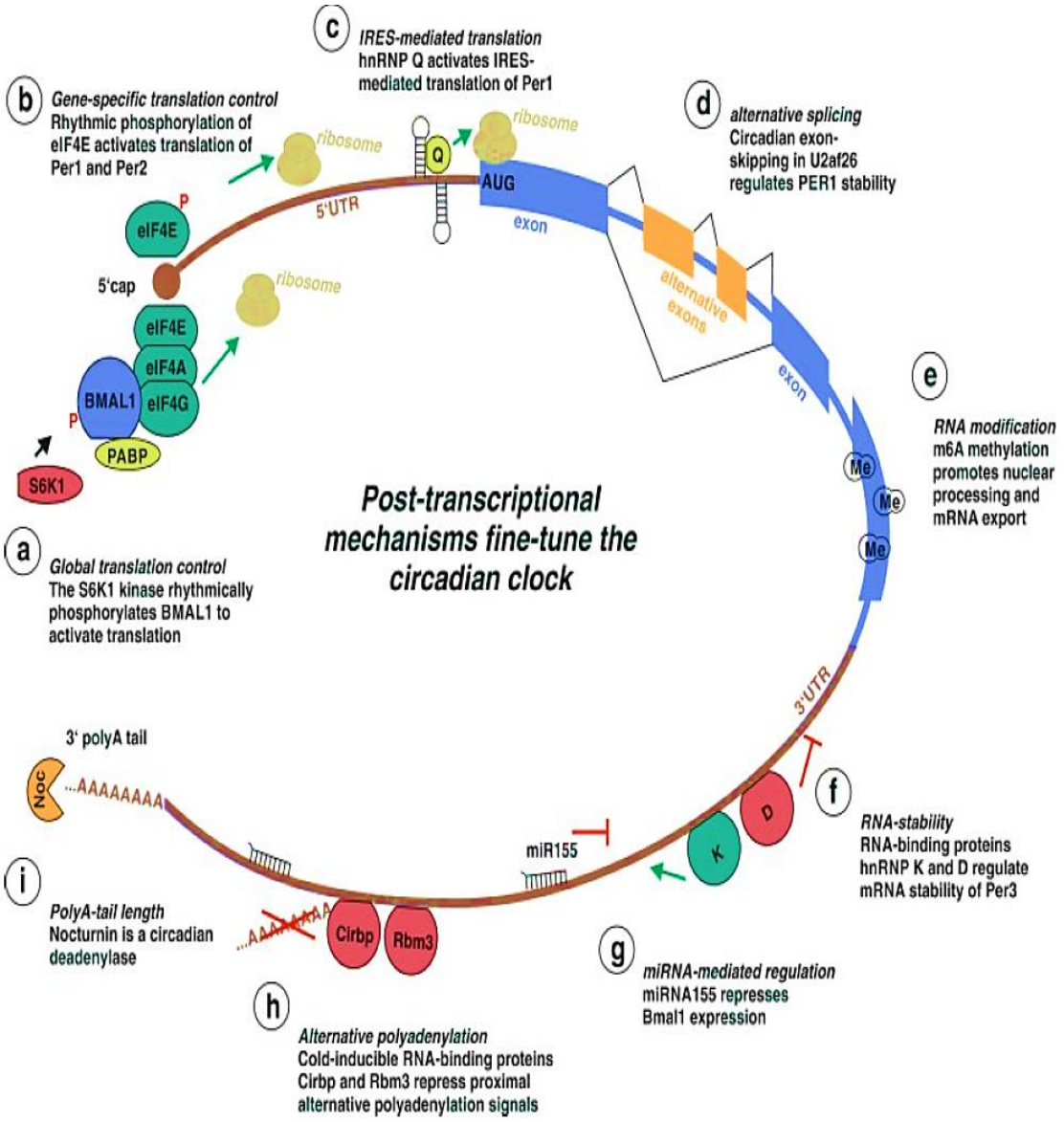


Fig.8. Post-transcriptional mechanisms acting on the (pre) mRNA to regulate the mammalian circadian clock.

Preubner and Hayed, 2015

Post-translational mechanisms

Among the most compelling evidences for the presence of complimentary oscillatory mechanisms based solely on phosphorylations of clock proteins emerged from lower organisms as the cyanobacteria (Leloup, 2009). The significance of similar Post-translational oscillators (PTO) which supports the TTLF oscillator was later established in mammals. Overexpression experiments involving PER2 proved that rather than the absolute quantity, the dynamic protein oscillation is crucial for behavioural rhythmicity (Chen et al., 2009) Mainly, Posttranscriptional modifications (PTMs) allow slow and progressive maturation of negative transcription factors (PER and CRY) into transcriptional repressors. This builds in a delay between transcriptional activation and repression contributing to the circa 24h endogenous period of the molecular clock. (Mehra et al., 2009). PTMs mainly regulate the subcellular localization and stability of PER:CRY complexes through the opposing activity of kinases, phosphatases, and ubiquitin E3 ligases (Gustafson and Partch, 2015).

Dynamic phosphorylations regulate the temporal abundance and turnover of clock proteins

Phosphorylation is most widely addressed among PTMs that regulate clock function. CLOCK, BMAL1, CRY as well as PER undergo rhythmic changes mediated by specific Kinases, mainly Casein Kinases (CKs), Glycogen Synthase Kinases (GSKs), Adenosine Monophosphate-activated protein Kinases (AMPKs), Protein kinase A (PKA) Calcium/calmodulin-dependent protein kinase (CAMK) etc. Among the Protein Phosphatases (PP) that regulate clock protein functions are PP1, PP2A, PP4 and PP5 (Reischl and Kramer, 2011). The figure given below (Fig.9) summarizes the role of phosphatases and dephosphatases in regulating clock functions.

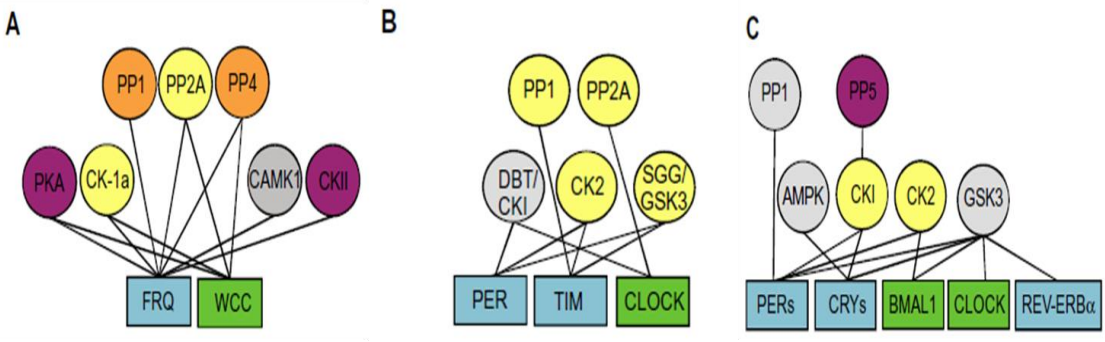


Fig. 9. Schematic representation of the main components, kinases and phosphatases and their interaction in the molecular circadian clock of Neurospora, Drosophila and mammals.
Transcriptional repressors in the molecular clockwork are depicted in light blue, transcriptional activators in green. Kinases and phosphatases are coloured according to their effect on circadian period. When inhibition of kinase/phosphatase activity causes period lengthening, shortening or arrhythmia, the kinase is depicted in yellow, orange or purple respectively. With no or unclear effect on period the kinases/phosphatases are depicted in grey. Lines between components represent interactions between kinase/phosphatase and its substrates. (A) Neurospora): PP1 – protein phosphatase 1, PP2A – protein phosphatase 2A, PP4 – protein phosphatase 4, PKA – protein kinase A, CK1-a – casein kinase Ia, CAMK-1 – calcium/calmodulin-dependent protein kinase, CKII – casein kinase 2, FRQ – frequency, WCC – white collar complex. (B) (Drosophila): PP1– protein phosphatase 1, PP2A – protein

phosphatase 2A, DBT/CKI – DOUBLETIME/casein kinase I, CK2 – casein kinase 2, SGG/GSK-3 – SHAGGY/glycogen synthase kinase 3, PER – PERIOD, TIM – TIMELESS, CLK – CLOCK. (C) (mammals): PP1 –protein phosphatase 1, PP5 – protein phosphatase 5, AMPK – adenosine monophosphate- activated protein kinase, CKI – casein kinase I, CK2 – casein kinase 2, GSK-3 – glycogen synthase kinase 3, PERs – PERIOD proteins 1-3, CRYs – CRY proteins 1 and 2, BMAL1 – brain and muscle aryl hydrocarbon receptor nuclear translocator (ARNT)-like 1),CLOCK – circadian locomotor output cycles kaput, REVERB-a – nuclear receptor subfamily 1, group D, member 1.

Reischl and Kramer, 2011

The rhythmic abundance of PER is regulated by its time specific interaction with CK1Ɛ/∂, the protein is progressively phosphorylated and targeted for proteasome dependent degradation by F-box proteins β-TrCP1/2 that are part of the E3 ubiquitin ligase complex (Ohsaki et al., 2008). BMAL1 and CRY 1 are also substrates for CKI activity. The Neurospora CK1 homologue, CK1-a phosphorylates and destabilizes the transcriptional repressor FRQ. In drosophila, DOUBLETIME (DBT, orthologue to vertebrate CKIƐ) also phosphorylates dPER for recognition by the F-box protein SLIMB and subsequent ubiquitination and proteasomal degradation (Reischl and Kramer, 2011). In mammals PER2 exists in two functionally different phosphorylation states, one primarily mediating proteasomal degradation and the other nuclear retention and stability (Vanselow et al., 2006) (Fig.10). Phosphorylation at Ser662 residue by an unknown kinase triggers phosphorylation at a conserved downstream phospho-acceptor cluster by CKIƐ/∂ leading to nuclear retention and stabilization. Interaction of PER2 with CRY1/2 favors protein stability (Yagita et al., 2002) possibly by mediating phosphorylation at this Ser166 residue. Inability to phosphorylate this priming site, either due to a point mutation on PER2 or a defective kinase activity may trigger alternate phosphorylation events that specify premature nuclear exit and proteasomal degradation of PER2.

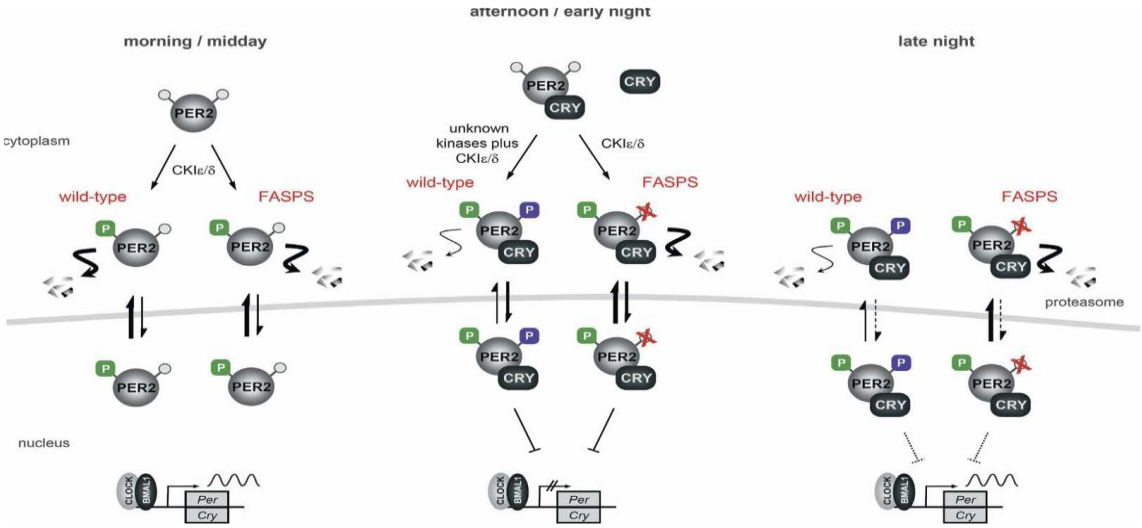


Fig.10. Model for the differential effects of PER2 phosphorylation on circadian oscillations.
PER2 contains at least two functionally different sets of phosphorylation sites—one primarily mediating proteasomal degradation (green), the other nuclear retention (purple). In FASPS-PER2 (*right* side of the panels), the latter cannot be phosphorylated because Ser 659 is mutated to glycine. The region

responsible for nuclear retention cannot be phosphorylated (red crosses), leading to premature nuclear export of the PER2–CRY complex, and thus to an earlier cytosolic degradation and to a faster circadian cycle.

Vanselow et al., 2006

Ser662G mutation is also associated with enhanced transcriptional repression (Xu et al., 2007) In Humans Familial Advanced Sleep Phase Syndrome (FASPS), the Mendelian inherited defective circadian phenotype is associated with the above mentioned defects on PER2 protein or the Kinase activity itself. The affected individuals experience a significantly short endogenous clock in comparison to the unaffected population (Chong et al., 2012). Chiu et al., 2011 has proposed a 'Time delay phospho circuit' model of dPER where phosphorylation at Ser596 by NMO/NLK kinase stabilizes the protein against the destabilizing DBT mediated phosphorylation at an alternate SLIMB recognition site at Ser47. The discrepancy in the circadian phenotypes observed in CKI *tau* mutation models can be explained by the Ser662 dependent fate of phosphorylation events. Moreover the model gains more significance in the light of differential effect of CRY 1 and 2 on PER2 stability as Cry1−/− and Cry2−/− mutant mice show short and long circadian phenotypes, respectively (Vanselow et al., 2006). Altering the preference of phosphorylation sites on PER2, either the FASP or βTrCP1 phospho binding sites regulates the fraction of stable and unstable PER2. This dual kinase 'phosphoswitch' is thought to communicate and integrate various environmental stimuli, such as temperature variations to the molecular clock, making it 'temperature compensated'. Effective period lengthening at higher temperature is attributed to increased FASP versus βTrCP binding site phosphorylation, as the FASP priming kinase is predicted to be more temperature sensitive than CKI (Zhou et al., 2015).

Ubiquitination and degradation of CRY1 and CRY2 *via* the Skp1-Cul1-FBXL3 (SCF- FBXL3) ubiquitin ligase complex is mediated by AMPK dependent phosphorylation. CRY2 is also phosphorylated at the priming site, Ser557 by DYRK1A kinase followed by subsequent phosphorylation at Ser557 by GSK-3b leading to FBXl21 mediated degradation. Unlike Fbxl3, Fbxl21 expression is rhythmic with a peak at the end of a subjective day. Fbxl21 temporally organize the stable accumulation of CRY in the cytosol and also due to its higher affinity for CRY, binds and counteracts the destabilizing effect of Fbxl3 in the nucleus until its levels decreases (Hirano et al., 2013; Yoo et al., 2013). The stability of REV-ERBα, BMAL1 etc. are also regulated by time specific phosphorylation and degradation. GSK-3β is responsible for the phosphorylation and stabilization of REV-ERB, whereas ubiquitination and degradation is the consequence for BMAL1. The proteolysis of BMAL1 coincides with the time of highest transcriptional activity, while transcriptional repression is associated with stabilized BMAL1 due to PKCϒ mediated deubiquitination (Stojkovic et al., 2014). The deubiquitinating enzyme, ubiquitin specific protease 2 (USP 2) is activated by CLOCK/BMAL1 promoter binding and is also critical for both the endogenous and entraining functions of the molecular clock (Yang et al., 2014).

Other significant PTMs and Epigenetic factors regulating the clock

PTMs involving small ubiquitin like modifiers (SUMO) is well reported, mBMAL1 is rhythmically SUMOylated in a CLOCK-dependent manner and is important for potentiating transcription (Mehra et al., 2009). SUMOylation of CLOCK is also associated with cell growth and proliferation *via* enhanced transcriptional activity of Eostrogen Receptorα (ERα) (Li et al., 2013). Acetylation has been a part of the 'histone code' regulating the dynamic changes in gene expression. CLOCK is known to have Histone acetyl transferase activity (HAT) (Doi et al., 2006) and it also acetylates and activates BMAL1. CLOCK acetyltransferase activity is counterbalanced by the rhythmic binding of the NAD+-dependent deacetylase SIRT1. As the TF heterodimer modulates the levels of NAD+ in the cell through transactivation of Nicotinamide phosphorybosyl transferase (NAMPT) the rate-limiting enzyme in NAD+ biosynthesis, SIRT1 binding is thought to integrate information on energy metabolism into the clock (Rey and Reddy, 2013). This interlink is further reinforced by the recruitment of poly (ADP) ribose polymerases (PARPs) which functionally interacts with SIRT1. The activation of PARP1 following cellular stress depletes cellular NAD+ thereby reducing SIRT1 activity and leading to cell death. PARP 1 also poly (ADP) ribosylate CLOCK modulating the binding of circadian transcription factors (Masri et al., 2012) CLOCK-BMAL1 bound SIRT1 also mediates deacetylation and degradation of PER2 (Chong et al., 2012). Epigenetic mechanisms involving chromatin remodelling is also hence an integral component of clock function (Fig.11).

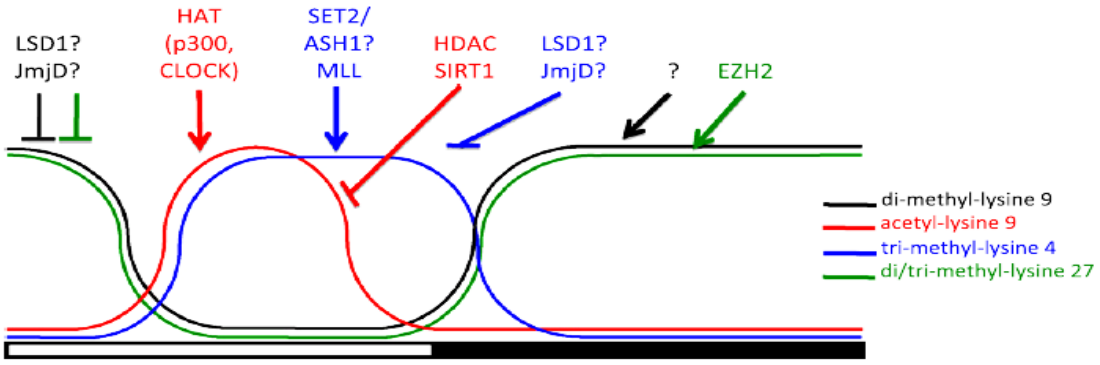
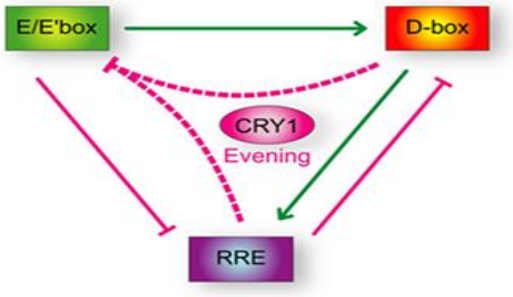


Fig.11. Model of dynamic chromatin transitions at circadian genes over the 24 h day (day, open bar at bottom; night, closed bar at bottom). Activation of BMAL1 and CLOCK regulated genes coincides with acetylation of histone residues by HATs, and tri-methylation of lysine 4 of histone H3 by Set2/Ash1 (via a WDR5 adapter) or MLL, and the removal of repressive histone methylations by LSD1 or JmjD-domain proteins. These activities are counterbalanced by the recruitment of HDACs or SIRT1, and LSD1 and JmjDdomain proteins, respectively. Similar scenarios may occur at RORa/REV-ERBa and DBP regulated circadian target genes as well. Indeed, REV-ERBa was recently identify as the main anchor point for HDAC3

Ripperger and Merrow, 2011

Temporal phase relationship between clock genes- the clock promoter elements.

The temporal transcriptional control and phasing of clock transcripts is controlled by mainly 3 clock-controlled DNA elements. The morning time- E/E'box [CACGT(G/T)] and the day time D box [TTA(T/C)GTAA] regulatory elements mediate the expression of *Reverb, Per1,2* and *Per3*. The Rev-ErbA/ROR response element (RRE-AGGTCA) is important for the night time expression of *Bmal1*, antiphasic to *Per2* oscillations (Ueda et al., 2005) The 'combinatorial regulation' involving E/E' and D box along with the newly discovered intronic RREs confer phase delay in *Cry1* expression at CT12, much later than the morning time genes and intermediate to the day time and night time elements (Fig.12). This delayed negative feedback exerted by the otherwise morning inhibitor *Cry1* expressed in evening, plays an important role in keeping the biological clock on time (Ukai-Tadenuma et al., 2010).



Ukai-Tadenuma et al., 2010

Fig.12. A minimal circuit model for the mammalian circadian transcriptional network.
Morning (E/E') box and (D-box) and night (RRE) DNA elements lie at the heart of the circadian clock Solid lines indicate activating (green) and inhibitory (pink) interactions between clock gene types. The dotted lines indicate how *Cry 1* represses morning expression through the combined action of day and night sequences

PER2 the multifaceted molecule

The *Per* 2 gene has been described as 'speciation gene' (Coyen, 1992) considering its role in regulating animal behavior that has direct relevance to courtship and mating (Tauber et al., 2003), leading to Sympatric speciation. Its utility as a plausible phylogenetic marker was also explored in Dipterans (Bauzer et al., 2002) and Lepidopterans (Regier et al., 1998). In addition to being the first clock gene to be identified, the direct link between *Per2* locus and endogenous period, demonstrated by 'short', 'long' and 'null' mutants of drosophila is a classic example of a single gene affecting a behavioral phenotype (Konopka and Benzer, 1971). It is the turn-over rate of PER2 that encodes the period of an endogenous day (Fig.13) the FASP (Familial Advanced Sleep Phase) syndrome among humans being an evidence to this particular function. It has been proposed that *Per2* may be involved in an evening oscillator which tracks dusk (Daan et al., 2001) it is hence speculated that a mutation or polymorphism that perturbs the evening oscillator may manifest itself in increased 'morningness'. Adding on more confidence to this speculation, a polymorphism in the 5'UTR of *Per2*, named as C111G is found to modulate diurnal preferences in humans, with the G allele

being more prevalent in subjects with extreme morning preference. The 111G and C alleles have different RNA secondary structures altering their translational efficiency. Based on computer simulations 111G allele is translated effectively leading to faster PER2 accumulation and clearance resulting in a shorter internal time (Carpen et al., 2005). Single nucleotide polymorphisms (SNPs) in the 3'UTR region (PER2-2221 A/G) as well as in the protein coding domains (PER2 G3853 A- 1244th amino acid position) of PER2 has also been associated with extreme morning preference in Korean adults (Lee et al., 2011; Lee et al., 2015). A very interesting report has correlated SNPs in clock genes as a mark of selection pressure induced by migration of human population to different geographic latitudes with varying photoperiods (Forni et al., 2014). Among other clock genes, *Per2* along with *Vip* is indispensable for maintaining the structural and functional integrity of SCN (Branacaccio et al., 2014). Photic induction of *Per2* via its CRE also facilitates phase shifts in behavioral rhythms such that they are synchronized to the natural photic cycles and hence contributes to the entraining properties of biological clock. As discussed previously the stability status of PER2 is also significant in conveying temperature variations to the clock (Zhou et al., 2015).

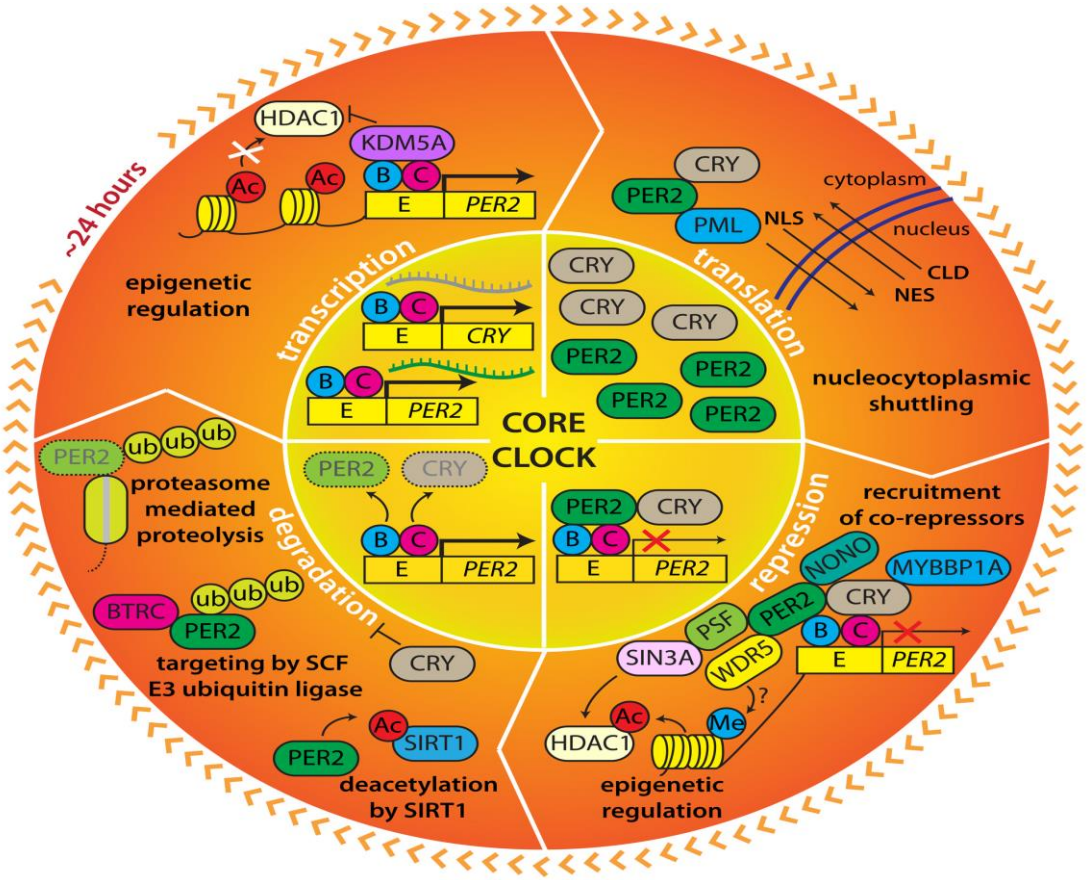


Fig.13. Regulation of PER2 intracellular dynamics plays a crucial role in core clock function

The core clock (inner circle, yellow) consists of a negative feedback loop driven by the transcription, translation, repression, and degradation of core clock components PER2 and CRY, which together drive the circadian rhythms of a cell. The outer circle in orange illustrates molecular mechanisms pertaining to PER2 intracellular dynamics, which in turn affect core clock events. Together, these processes take ~24

hours to complete. Abbreviations: Ac-acetyl group, B-BMAL1, C-CLOCK, E-E-box promoter elements, ub-ubiquitin. HDAC1- histone deacetylase 1. KDM5A-lysine (K)- specific demethylase 5A. PML-promyelocytic leukaemia protein. NLS-nuclear localization sequences, BTRC-beta-transducin repeat containing E3 ubiquitin protein ligase, NONO non-POU domain containing, octamer-binding, SIN3A-SIN3 transcription regulator homolog A, SIRT1-Sirtuin 1, CLD-cytoplasmic localization domain, NES-nuclear export sequence, PSF-Polypyrimidine tract binding protein-associated splicing factor, WDR5-WD repeat domain 5, MYBBP1A-MYB binding protein

Chong, et al., 2012

Per2 was the only mRNA that was commonly identified as being rhythmic among microarray data analysis from both SCN and liver (Kojima and Green, 2014). Acetylation of PER2 by SIRT1 communicates the status of energy metabolism in peripheral clocks to the SCN. Temporally controlled interactome of PER2 acts as coactivators or corepressors in regulating metabolic pathways (Ripperger and Albrecht, 2012). Also, through its regulation of p53 and *cmyc* expression loss of *Per2* is associated with cell cycle regulation and cancer (Fu et al., 2002) (Fig.14). Neurologically, PER2 is involved in the reward and addiction mechanism (Jagota and Reddy, 2007) sleep homeostasis depression vulnerability, neurodegenerative disorders (Reddy and Jagota, 2014; Mattam and Jagota, 2015) synaptic plasticity, learning etc.

As a multifaceted molecule *Per2* is hence an ideal candidate for comparative studies at the behavioral level, offering both phylogenetic as well as mechanistic insights into the functioning of the molecular clock.

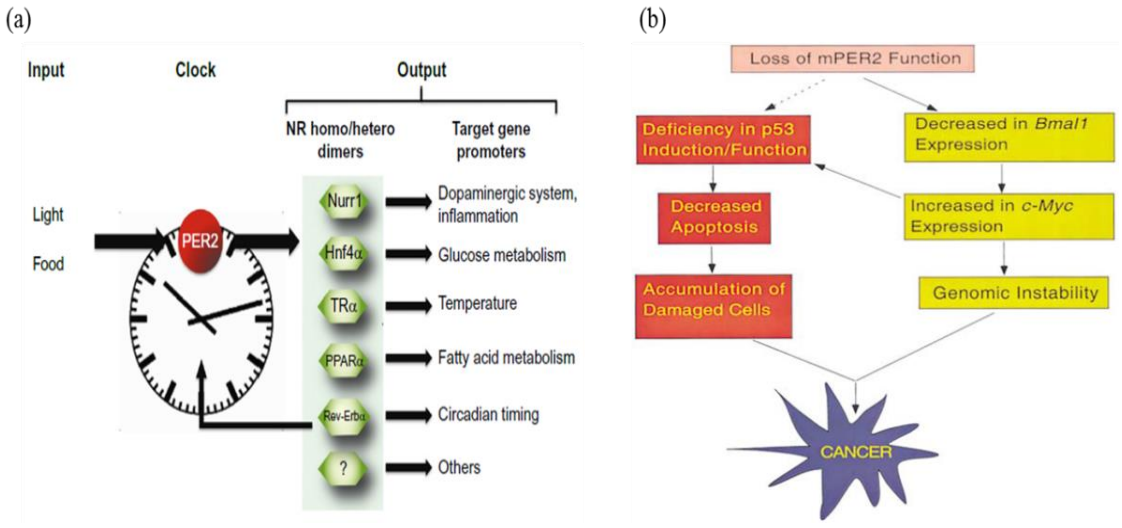


Fig. 14. (a) Transmission of circadian clock information to nuclear receptor-target genes by PER2. PER2 mediates primary output from the molecular oscillator. By direct interaction with the nuclear receptor (NR) homo- or heterodimers, it can affect the corresponding target gene promoters and metabolic or physiological processes. The activity of PER2 may be modulated by the input (e.g., light or food) to the circadian oscillator. (b) A Proposed Model for the Role of *mPer2* in tumour suppression. The solid lines indicate the pathways that have been demonstrated. The dashed line indicates a regulatory pathway(s) that is still not fully understood

Ripperger and Albrecht, 2012; Fu et al., 2002

Animal models in Chronobiology

Temporal distribution of behavior in response to time dependent environmental stressors like food availability, temperature and predation has led to the evolution of alternate chronotypes, as diurnal, nocturnal, crepuscular etc. (De Coursey, 2004) (Fig.15). Events of inter-specific competition with the ectothermic diurnal reptiles during the Mezozoic had imposed prolonged episodes of nocturnality in early eutherian mammals (Gerkema et al., 2013) A nocturnal ancestry is hence proposed for all the extant mammals, and at least in Rodentia, diurnality evolved through independent events of secondary evolution (Roll et al., 2006) necessitating the study of many unrelated species to truly appreciate the mechanism favouring the shift. Inspite of the difference in photic experience most of the fundamental properties of the circadian timing system such as, diurnal rhythm in 2- deoxy glucose uptake by SCN, temporal profile of clock gene expression etc. are similar in both diurnal and nocturnal species (Koch et al., 2000). This has not only led to the wide usage of the conventional laboratory nocturnal rodent models for the study of circadian biology, but also to the assumption that diurnality rather than being a property within the SCN could be an interpretation of its output signals (Smale et al., 2003) A comparative approach to the understanding of 'temporal niche' preference using unconventional diurnal animal models like chipmunks (Abe et al., 1995), octodon degus (Lee, 2004), 13-stripped ground squirrel (Meijeir et al., 1989) has brought to light some significant differences between chronotypes with respect to photic response. Strikingly, some diurnal rodents (Abe et al., 1995; Krajnak et al., 1997; Schumann et al., 2006) like humans are sensitive to light throughout the biological day with little indication of a dead zone (Duffy and Czeisler, 2009) unlike nocturnal rodents where photic response is temporally gated.

Moreover, the present neuropsychiatric disease models based on nocturnal rodents are a major setback in the identification and development of animal models for psychiatric disorders that stems from circadian rhythm anomaly in humans (Kronfeld-Schor and Einet 2012; Ashkenazy-Frolinger et al., 2015) The necessity for diurnal animal models that would mimic human circadian disorders saw the emergence of *O. degus* as an advantageous model for the study of affective disorders (Ashkenazy-Frolinger et al., 2013), Alzheimer's disease (Tarragon et al., 2013) and melatonin research (Lee et al., 2009). Cognitive impairment of *O. degus* is comparable to that of humans, and owing to its phylogenic status the structure of critical molecules such as AANAT and Aβ-APP of the degu show remarkable similarity to that of humans. Though it is not known how temporal signals from the SCN are translated into activity patterns, some basic differences between diurnal and nocturnal animals, like the rest activity timing with respect to melatonin rhythms and the masking effect of light emphasizes the importance of diurnal rodents for the study and management of photoperiod related affective disorders (Ashkenazy-Frolinger et al., 2013) According to Rissman (2004), comparative systems biology therefore confers scope for understanding the molecular and genetic mechanism underlying the variability in

physiological processes, offering solutions for basic as well as applied clinical problems.

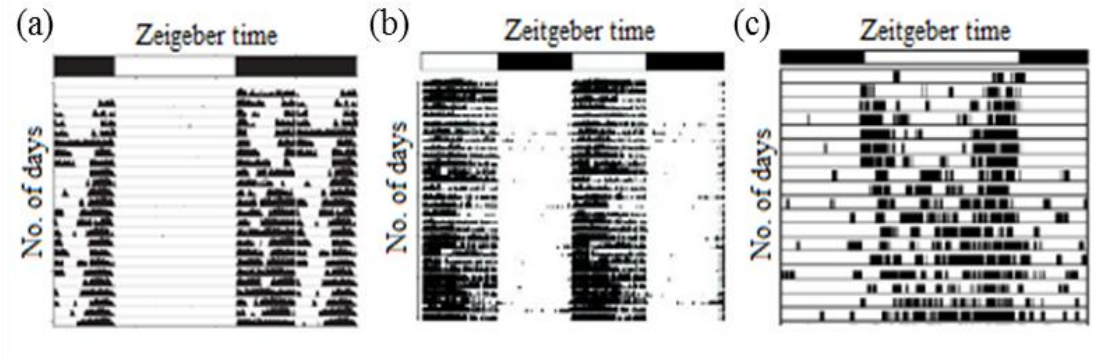


Fig15. Representative actograms showing locomotory rhythms of a (a) nocturnal (b) diurnal and (c) crepuscular organism.

Reddy and Jagota, 2014; Mammen and Jagota, 2011; Forster-Helfrich, 2005

The south Indian palm squirrel, *Funambulus palmarum* as a promising diurnal model for circadian biology

Funambulus palmarum commonly called as the south Indian palm squirrel is a species of rodent belonging to the family Scuridae, they are naturally day active and unlike other species do not hibernate in the winter. Through behavioral and neurochemical approaches our laboratory has proposed the utility of this animal as an ideal diurnal model for the study of circadian biology (Fig.16). *F. palmarum* shows definitive diurnal activity under both endogenous (DD and LL) and entrainable conditions (LD) which persists in the presence of a running wheel (Mammen and Jagota, 2011). While a compromise in diurnality has been observed with *O. degus* and *Arvicanthis* under conditions of varying temperature cycles, running wheel, social cues etc: (Redlin and Morowsky, 2004; Vivanco et al., 2010) the locomotory activity of our model confines to the subjective day even when food availability is restricted to the night time (Mammen and Jagota, unpublished). Apart from its diurnal features in 'circadian neuroanatomy' the SCN of this animal shows structural heterogeneity with respect to VIP and AVP immunoreactivity. These neurotransmitters also showed a temporal pattern of expression within the SCN. Importantly, uncoupling of the multi-oscillator was demonstrable in terms of 'splitting' of locomotory rhythms (Mammen and Jagota, 2011) as well as phase reversal of VIP-ir in presence of constant light (LL) (Mammen and Jagota, 2011).

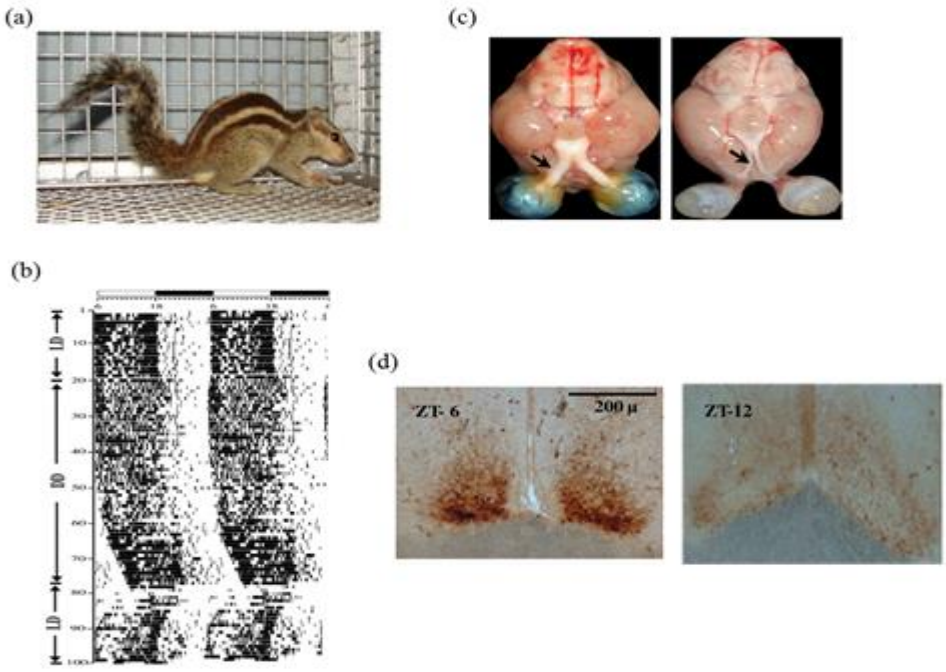


Fig. 16. (a) *Funambulus.palmarum* (South Indian three stripped palm squirrel) housed in a cage (b) Actogram showing the locomotory activity confined to the light phase of an LD cycle and to the subjective day under DD (c) The circadian neuro anatomy of the model (left panel) in comparison to nocturnal rat (d) VIP-ir peaks at ZT 6, at the SCN core region (left) whereas AVP-ir peaks at ZT 12 delineating the shell region of SCN

Mammen and Jagota, 2011

Our laboratory has also investigated the serotonin metabolome of SCN and pineal in *F.palmarum* by Reverse Phase HPLC analysis of daily temporal profile of various compounds involved in Serotonin metabolism, including Melatonin. Rapid onset of Melatonin synthesis at night associated with posttranscriptional regulation of AANAT is opined to be characteristic of humans. Interestingly, the Melatonin surge in the pineal of *F.palmarum* occurs with little latency at night (Mammen and Jagota, 2011) in comparison to the delayed kinetics of melatonin synthesis in nocturnal rats (Reddy and Jagota, 2015). This observation is supported by a similar pattern of Melatonin production in *O. degus,* where a distinct N-terminus region of *Aanat* mRNA is linked to the posttranscriptional regulation of its protein production (Lee et al., 2009). However, this study also states that it is the phylogenetic position of the organism rather than its diurnality which contributes to a human like pattern of melatonin synthesis and regulation. *F. palmarum* therefore holds the dual advantage of being phylogenetically advanced and behaviourally diurnal in offering better scope to a comparative approach aimed at understanding the varied expressions of the biological clock.

With this literature background the objectives of the present work were designed with the aim of characterizing the molecular elements of the central clock in *F. palmarum*. This would not only validate the utility of our animal model but also may provide future leads to the understanding of differentiating factors that make up a diurnal clock.

Objectives

1. **Cloning and Characterization of *Per2* CDS from the SCN of *F. palmarum***

2. **(a) Cloning and Localization of SCN expressed Vasoactive Intestinal Peptide (*Vip)***

   (b) Localization of *Per2* transcripts in SCN using *Vip* as a marker for 'core'

3. **Effect of various photoperiods on Clock gene expression . *Per1, Per2, Per3, Cry1, Cry2* and *Bmal1***

4. **Characterization of the 1st intronic sequence of *F. palmarum Per2*: a site for potential regulatory elements & *Per2* as a possible phylogenetic marker**

Materials and Methods

Entrainment of Animals

Entrainment and maintenance of animals was done as per the methodologies established previously in our laboratory (Mammen and Jagota, 2011). Adult male 3 striped South Indian Palm Squirrel, *Funambulus palmarum* were maintained in 12 h light and 12 h dark condition (LD 12:12; lights on at 6am, off at 6pm), at room temperature 24±2°C with relative humidity of 55±6 for 2 weeks . Animals were individually housed in cages equipped with infra-red motion detection sensors. Food and water were provided *ad libitum* The gross locomotor activity was recorded using a chronobiology kit (Stansford, USA) under LD condition for 2-3 weeks . For recording the endogenous rhythms, after entraining the animals to an LD regime, the timer were set to constant dark (DD) and the actograms were recorded until the activity rhythms showed a stable free running pattern (Mammen and Jagota, 2011). Animal handling was done using a squirrel handling cone (Koprowsky, 2002) which was modified according to the morphological dimensions of this model. For experiments in 'constant darkness' (DD) animals were initially entrained to an LD schedule as described and further kept under DD allowing the endogenous clock to free-run. Animal handling and maintenance in dark were done under dim red light. All the experiments were done as per Institutional Animal Ethics regulations.

C57BL6 mouse, *Mesocricetus auratus* (Syrian hamster) and *Cavia porcellus* (Guinea pig) were used for comparative studies involving SCN genomic DNA (gDNA) These animals were maintained in separate cages in LD conditions as explained earlier.

For qRT-PCR analysis animals were sacrificed at 4 time points, ZT0, ZT6, ZT12 and ZT18 (n=4 for each time point) and the brain from individual animals were carefully dissected out and snap frozen in liquid nitrogen. The tissues were stored in -80 till further use. For objectives 1 and 2, ZT12 brain was used considering the expression peak of *Period2* transcript at this time point in other rodents (Reddy and Jagota, 2014; Mattam and Jagota, 2015) Brain tissue was prefixed in 4% paraformaldehyde and 20% sucrose solution prior to cryosectioning for the In-situ hybridization experiments.

Primer designing

All the primers used for the study were designed across conserved domains of the respective transcripts. These oligonucleotides were projected using an alignment of already reported nucleotide sequences from different mammalian species using the CLUSTALW programme. For the analysis of *F. palmarum clock genes, Per1, Per2, Per3, Cry1, Cry2 and Bmal1* NCBI Genbank data from the following organisms were utilized, with the respective NCBI Accession No. for each gene indicated in brackets. *Mus musculus* (NM_011065.4, NM_011066, NM_011067.2, NM_007771.3, NM_009963.4, NM_007489.3), *Rattus norvegicus* (NM_001034125.1, NM_031678, NM_023978.2, NM_198750.2, NM_133405.1, NM_024362.2), *Homo sapiens* (NM_002616.2, NM_022817, NM_016831.1, NM_004075.3, NM_021117.3,

NM_001178.4), *Macaca mullata* (XM_594471.3, XM_001086845, XM_002802160.1, NM_001194159.1, XM_001113162.2) and *Bos Taurus* (XM_594471.3, NM_001192317, XM_002701561.1, NM_001105415.1, XM_002693564.1). Additionally for *Per 2*, nucleotide information from other rodent species as *Arvicanthus niloticus*, *Octodon degus* and *Mesocricetus auratus* (AY817663.1, EU590918.1, XM_013122132.1) were used for CLUSTAL alignment. For PCR amplification and cloning of Vasoastive Intestinal Peptide (Vip) the following sequences were used from NCBI, *Mus musculus* (NM_011702.2), *Rattus norvegicus* (NM_053991.1) and *Homo sapiens* (NM_003381.3). We have also designed primers for *Gapdh*, which would act as an endogenous control in gene expression studies. Alignment between gene paralogues was also done inorder to ensure primer specificity. Primer designing was supported by basic on-line tools as IDT-Oligo Analyzer, OligoCalc etc. and their specificity was ensured using the BLAST algorithm. Primers across gDNA were designed using NCBI-genomic database information on *Mus musculus* (NC_000067), *Rattus norvegicus* (NC_005108) and *Homo sapiens* (NC_000002).

Total RNA extraction and cDNA synthesis (Chomczynski and Sacchi, 1987)

The SCN was carefully punched out from 500µm coronal slices of specific brain tissues, mRNA isolation was performed in 500µl of TRI reagent as per the standard extraction protocol of the manufacturer (TRI, Sigma) The RNA pellet was dissolved in 20µl RNase free DEPC water, mass quantity was measured by Nanodrop (ThermoFischer), further 1µg of RNA was used for cDNA synthesis (Superscript III, Invitrogen) with oligodT primers priming the reverse transcriptase reaction under the standard reaction conditions proposed by the manufacturer.

Genomic DNA (gDNA) isolation by organic extraction

gDNA was extracted from SCN of *F.palmarum* using organic extraction method (Strauss,1998). The tissue was quickly and gently homogenized in 600µl of digestion buffer containing 100mM NaCl, 10mM Tris.Cl (pH8), 25mM EDTA (pH8), 0.5% SDS and Proteinase K at 0.1mg/ml concentration added only at the time of experiment. Homogenate was incubated with gentle agitation at 50° C for overnight. Further, equal volume of Phenol-Chloroform-Isoamyl alcohol mixture was added, gently mixed and centrifuged at 10,000rpm for 15min. at room temperature (RT). The clear top layer was separated into a fresh Eppendorf to which 5M Ammonium acetate to a final concentration of 2.5M and one volume of 100% ethanol were added, the reaction set up was incubated on ice for 8-9h. After centrifugation at 10000 rpm for 25 min at RT the resulting pellet was again washed by centrifugation in 70% ethanol, air dried and resuspended in 300µl TE buffer. Dissolution was done by incubating the sample at 37° C for 2h followed by overnight incubation at 4° C. The quality of the gDNA was confirmed by spectrophotometry as well as by Agarose gel electrophoresis. The method was

followed for gDNA extraction from the SCN of *Misocricetus auratus* and *Cavia porcellus* as well.

Semiquantitative RT-PCR (Metzenberg, 2007)

The general PCR reactions were performed using Taq DNA polymerase (Dream Taq, Thermo Fischer) with 0.025-0.05µg RNA equivalent of cDNA as template (Table 1). Long Range PCR reactions with cDNA or gDNA as template were standardized using Elongase Reaction Mix (Invitrogen) and KOD-FX enzyme (Toyobo), Tables 2 and 3 respectively. The primer information mentioned in Table6 was exclusively used with gDNA template. Table 7 and Table 8 enlists the primers designed for cDNA template in conventional and long range PCR reactions, respectively. Gel electrophoresis on 1or 1.5% Agarose in 0.5x Tris Borate EDTA (TBE) buffer confirmed the presence of amplicons, as compared against standard Gene ladders. DNA was gel extracted using commercial kits (Quiagen) and the eluates were send for analysis by sequencing (CDFD, Hyderabad).

The PCR reaction set up, in the order of reagents added and their respective cycling conditions were as follow

Table1.

A conventional PCR set up		
1.	Sterile MilliQ H_2O	To make upto 25µl
2.	10x Dream taq Buffer	2.5µl
3.	25mM dNTP mix	0.25µl
4.	Sense Primer (5pm/µl)	0.5µl
5.	Antisense Primer (5pm/µl)	0.5µl
6.	cDNA template	0.025-0.05µg RNA eq.cDNA
7.	Dream Taq Enzyme	2.5U (0.25µl)

Initial denaturation: 95°C,2 min.

Cycle denaturation: 95°C, 30 sec.
Annealing: 58/60°C, 30 sec.
Extention: 72°C, 40-60 sec.
} 35 cycles

Final extension: 72°C, 7-20 min.

Long Range PCR reaction set up:

Table2.

Elongase PCR reaction		
1.	Sterile MilliQ H_2O	To make upto 50µl
2.	Buffer B	19µl
3.	10mM dNTP mix	1µl
4.	Sense primer (5pm/µl)	1µl
5.	Antisense primer (5pm/µl)	1µl
6.	cDNA template	0.05µg RNA eq. cDNA 2µl
7.	Elongase enzyme	1µl

Initial denaturation: 94°C, 30 sec.
Cycle denaturation: 94°C, 30 sec. } 35 cycles
Annealing: 58°C, 35 sec.
Extension: 68°C, 5 min.

Table 3.

KOD-FX PCR reaction		
1.	Sterile MilliQ H_2O	To make upto 50µl
2.	2x PCR buffer	25µl
3.	2mM dNTP	10µl
4.	Sense Primer (5pm/µl)	2µl
5.	Antisense Primer (5pm/µl)	2µl
6.	gDNA template	40-80ng
7.	KOD FX enzyme	1µl

Initial denaturation: 94°C, 2 min.
Cycle denaturation: 98°C, 10 sec.
Annealing: 60°C, 35 sec. } 35 cycles
Extension: 68°C, 10 min.

TOP-10 DH5α Competent cell preparation

Single colony of TOP-10 cells grown overnight on LB-agar plates was used to inoculate 150ml LB media, incubated overnight at 37℃ at 140rpm. OD was measured every 30min/1h until the desired OD of 0.3 was attained (approx. 2.5h) The cells were pelleted down by centrifugation at 4000rpm for 15min.at 4°C and re-suspended gently

in 40 ml of ice cold, sterile filtered CCMB80 buffer, pH 6.4 (10mM KOAc, 80mM $CaCl_2.2H_2O$, 20mM $MnCl_2.6H_2O$, 10% glycerol) and incubated with infrequent gentle stirring for 1h on ice. Centrifugation was performed at 4000 rpm for 15 min at 4℃ and the cell pellet was again re-suspended in 5ml of ice cold CCMB80 buffer. The chilled buffer was again added to the suspension such that the final OD would be 1.5-2, this was then aliquoted (150µl) into pre chilled cryo vials and stored in -80° C until use.

TOPO-TA/TA-XL (Invitrogen) cloning

The basic protocol is that of Sambrook and Russel, 2001. The gel extracted DNA (amplicon) is mixed with the TA vector, in the ratio 4:1 along with the salt solution as described by the manufacturer. This cloning mixture is incubated for 5min in RT and 25min on ice, followed by transformation of chemically competent TOP-10 DH5α cells. The transformation protocol involves 30 min incubation of the Cloning reaction mix with the competent cells on ice, and further heat shock at 42℃ for 45 sec and immediate chilling on ice for 5min. 200µl of SOC medium was added to the transformed cells and incubated for 1h at 37° C in an incubator shaker to allow expression of antibiotic resistance. Cells were plated on antibiotic containing LB-agarose plates and incubated overnight at 37° C, ampicillin (100µg/ml) was used for TOPO-TA cloning whereas kanamycin (50 µg/ml) for TOPO-XL cloning. Positive colonies were picked and grown overnight in 5ml LB media containing the specific antibiotic. Further screening was done by performing 'colony PCR' using 2µl of the overnight culture as template with both gene and vector specific primers. After gel electrophoresis the inoculum carrying transformed cells were used for plasmid extraction by commercial kit method (Qiagen). The plasmids were used for subsequent sequencing and subcloning experiments. A glycerol stock of the same was also stored and maintained in -80° C.

Quantitative-Real Time PCR

The qRT-PCR was carried out using the SYBR Green Master mix (Applied Biosystems) in ABI –Step one plus real time PCR machine (ABI) according to the manufacturer's protocol (Table 4 and Table 9). The identity of all the transcripts, *Per1*, *Per2*, *Per3*, *Cry 1* ,*Cry2* and *Bmal1* as well as the endogenous control, *Gapdh* was confirmed by sequencing. Comparitive Ct method was used to analyse the results (Mattam and Jagota, 2014). Dissociation curves for individual reactions confirmed the specificity of the primers used. Annealing and Extension was performed at 60° C for 1min as per the manufacturer's advice.

Table 4.

qRT-PCR reaction set up		
1.	Sterile MilliQ H_2O	2µl
2.	Sense Primer (10pm/µl)	1µl
3.	Antisense Primer (10pm/µl)	1µl
4.	2X SYBR green mix	5µl
5.	cDNA (1:1 diluted)	1µl
	Total volume	10µl

***In-situ* hybridization (ISH) for localization of *Per2* in the SCN**

Generation of cRNA Probes, PCR method (Urrutia et al.,1993; Leonard, 1995; David and Wedlich, 2001; Owens et al., 2006)

Gene specific primers with T3 (antisense) and T7 (sense) 'promoter' sequence as 5'over hangs (Table 11) were used for generating specific amplicons from TA plasmids carrying *F. palmarum Per2* and *Vip* Coding sequence (CDS). These PCR products were gel extracted and subjected to *'in-vitro'* transcription to produce Digoxigenin labelled cRNA probes (DIG-labelling kit,Roche). The reaction was incubated for 2h at 37° C. The labelled probe was stored in -80° C until use (Table 5).

Table 5

***In-vitro* transcription reaction.**		
1.	RNAse free MilliQ	To make upto 20µl
2.	PCR amplicon	150ng
3.	10x DIG labelling mix	2µl
4.	10x Transcription buffer	2µl
5.	RNA polymerase, T3/T7	2µl

Cryosectioning and hybridization (Okamura et al., 1995; Ban et al,1997; Hamada et al., 2001; de la Iglesia, 2007)

All the reagents and buffers were made in autoclaved RNAse fee DEPC water. Brain (ZT 12) was fixed in 4% paraformaldehyde in 0.1M phosphate buffer of pH 7.3, after 12 h it was transferred to 20% Sucrose made in phosphate buffer and kept for 24 h at 4° C. Sectioning was performed on the same day on a Leica Cryostat, hybridization was performed on free floating 40µm sections. Alternate sections, upto the entire extent

of SCN were probed for *Per2* and *Vip* messages, sense probes acted as negative control for specific transcript detection.

Sections were collected in 4x Saline Sodium Citrate buffer (SSC), and kept in 4° C until the next procedure. The 20x SSC stock solution contained 3M NaCl and 0.3M Sodium Citrate at pH 7.0. Pre-hybridization and permeabilization of sections was done as follows, Protease K treatment (0.1mg/ml in 10mM Tris-Hcl, 10mM EDTA, pH 7.4) for 5min. at 37° C, 4% paraformaldehyde in 0.1M Phosphate buffer for 5min and 0.25% acetic anhydride in 0.1 M Triethanolamine for 10 min at RT. The sections were then transferred to 4x SSC for 10 min at RT and later into the Hybridization buffer containing the sense or antisense probe in 1:1000 dilution at 60° C for 16h. The hybridization buffer consisted of 60% deionized Formamide , 10% Dextran sulfate, 10mM Tris-HCl at pH8.0, 1mM EDTA at pH8.0, 0.6M NaCl, 0.2% Laurylsarcocine, 200µg/ml tRNA, 1x Denhardt's solution (from 50x sock containing 1g BSA, 1g Ficoll 400, 1g Polyvinyl Pyrolidone in 100ml DEPC water), 0.25% Sodium Dodecyl Sulfate (SDS) and 10mM Dithiothreitol (DTT). Sections were washed in buffer containing 50% formamide and 50% 4x SSC for 5 min at 60° C, followed by 45 and15 min each at 60° C , every time a fresh buffer was used. Sections were transferred to 4x SSC and then treated with RNAse buffer, pH 8.0 containing RNAse A (100mg/ml) in 0.5M NaCl, 1mM Tris HCl and 1mM EDTA at 37°C for 30min. Sections were then treated twice for 15 min each with buffer containing 50% formamide and 50% SSC at 60° C. This is followed by treatment in 0.8x SSC for 30min at 60° C and 0.1M Tris-HCl,pH 8.0 /0.15M NaCl buffer at 4° C.

For Immunochemical detection the sections are given two changes of Tris-NaCl buffer, 15 min each. Sections are transferred to Blocking solution(1g Roche blocking agent in 100ml Tris-NaCl buffer) and incubated for 1h at RT. This was followed by overnight incubation at 4° C in alkaline phosphatase conjugated digoxigenin antibodies (1:1000 dilution in Tris-NaCl buffer). Next day, section were washed with Tris-NaCl buffer, twice for 5 min each. Later sections were incubated in buffer containing, 100mM Tris-HCl,pH 9.5 and 100mM NaCl for 5 min. followed by the solution containing the substrate, Nitroble Tetrazolium salt (NBT) and 5-brono-4-chloro-3 indoyl phosphate toluidium salt (BCIP) at room temperature until colour develops. Colour formation was stopped by adding 10mM Tris-HCl/1mM EDTA,pH 8.0 buffer. Sections were mounted on silane (Sigma) coated slides and observed under microscope for immunoreactivity.

All The Primer sequences that were designed for specific experiments are as follows:

Table 6**.** Primers for gDNA , PCR and sequencing

SI	PRIMER	PRIMER SEQUENCE
1.	F CON-E	GC GGT CAC GTT TTC CAC TAT GTG
2.	R113	A GCC ACT GCT CAT GTC CAC
3.	R intron1	AG TCT GCA AAA CGT GAA CGC C
4.	-140P2 intron1F	GTG TGT GGT GCT GGA ATT GAA GC
5.	Ham ConE +R1	CGA GTT GCA CCC ATT CTC ACG
6.	Gui ConE +R1	T ACG CAC CTT CCG TTC ACT CAC
7.	Per2 ConE +F1	GCA CTC ACT GAA ACT CTG CGA
8.	113 minus R1 Per2	G TGG TAG AGT GCC TTG CAT GC

Table 7. Primers used for *Per2* cloning and sequencing

SI	PRIMER	PRIMER SEQUENCE
1.	*M13F	GTA AAA CGA CGG CCA G
2.	*M13R	CAG GAA ACA GCT ATG AC
3.	F334	GAA CTC TGA GGG AGC TGA AGG TC
4.	F581	CT GTG TCC CTG GTT TCT GGG AAG
5.	F927	GCT GCT GGT AGA GAG GGT GCA
6.	F1232	GCT GGT CTA GCT TCA TCA ACC C
7.	F1517	ACT CTA GCC GGT GGA GAT CCG
8.	F2150	ATG GAC CTC ACC AAG GAG GTG
9.	F2626	GTA GAC ACC AAG CCC GAG TTC G
10.	F3040	GAC TAC ACA CCT GGC GCT TCT CG
11.	F3386	AGG ACC TCA TCT GGC TGC TGA TG
12.	R610	AC AGG ATC TTC CCA GAA
13.	R1232	GGT TGA TGA AGC TGG ACC AGC
14.	R1350	TGA ACT CTG GGC TGA AGG GTC TTC
15.	R2216	TCC GTG AAC CTC TGC AGG AAG C
16.	R2784	CTG AGC AAG TGC AGG GCT GTC
17.	R2972	TCC ATC TAC AGC TTGAGC TG
18.	R3316	TT CTC TGA GGA GTC AAT GCT TCC A
19.	Minus3R3’UTR	CT CTG TCT TCC AAG CAC CAT CTG
20.	Minus2R3’UTR	A ACA GAA GCC ATG TCA CAC CAG
21.	Minus1R3’UTR	CAA ACA TCG GGT GGA CCT ATG AG

Table 8. Primers used to generate full length *Per2* CDS

SI	PRIMER	PRIMER SEQUENCE
1.	F Start1(d)	GAG CCC ART RTG AAK GGA TAT GC
2.	3'UTR 944R	GCT GGG TGA GAG CTG ATG TCT

Table 9. Primers used for qRT PCR

SI	GENE	PRIMERS	PRIMER SEQUENCE
1.	*Gapdh*	F213	CAT CAC CAT CTT CCA GGA GCG
		R360	CGG AGA TGA TGA CCC TTT TGG C
2.	*Per1*	F798	CTT TGG CTC TAC TGC ACC ATC TCG
		R978	CAC CCG AAT CTT GGT CAC ATA TGG
		F2754	GCC TTG GTG CTC CCT AAC TAT
		R3088	CC TGA AAG TGC ATC CTG ATT GGA
3.	*Per2*	F1517	ACT CTA GCC GGT GGA GAT CCG
		R1821	CCT GAA GAG GAA GTG CGA GTT
		R2216	GCG GAA CTT ATG CAG GAA GCT
4.	*Per2v*	FE8/E9 1041	AGA TGT GGA TGA AAG GGC GGT C
		R3b insert	AG AGA CTC ATG CCA CAG GCA AG
5.	*Per3*	F681	GT CCT CCC CTA TTT GGT TCA TGT
		R861	CGG CGC TTT CAT CTA CTT CAA GAA
		F1649	CCT TGA AGA GAA AGT GCA TCT CCT
		R2086	GCT CCT GCT TTC TTG CTG AAG ATA
6	*Cry1*	F534	AC AAC CCC TCT GTC TGA TGA C
		R671	GCC TTT CCA AAC GTG TAA GTG C
7.	*Cry2*	F534	A TAA GCA CTT GGA ACG GAA GGC
		R671	T ACA AGT CCC ACA GGC GGT A
8.	*Bmal1*	F285	GGC TTC TTT GGT ACC AAC ATG CAA
		R471	AA TCC ATC TGC TGC CCT GAG AAT

Table 10. Primers used for *Vip* PCR and cloning

	PRIMER	PRIMER SEQUENCE
1.	*Vip* F202(d)	GAC TCT CTT CAG TGT GCT STT CTC
	Vip R657	A AGT CGG GAG AAT CTC CCT CAC

Table 11. Primers for cRNA probe generation with T3/T7 5'overhangs (in red)

	PRIMER	PRIMER SEQUENCE
1.	P2 T7+ F927	CCA AGC TTC TAA TAC GAC TCA CTA TAG GGA GA CT GCT GGT AGA GAG GGT GCA
	P2 T3+ R1238	CAG AGA TGC AAT TAA CCC TCA CTA AAG GGA GA GT TGA TGA AGC TGG ACC AGC
2.	P2 T7+ F1116	CCA AGC TTC TAA TAC GAC TCA CTA TAG GGA GA AG TGA CAG GCG CTT GAT GCT
	P2 T7F 1041 E8/E9	CCA AGC TTC TAA TAC GAC TCA CTA TAG GGA GA AGA TGT GGA TGA AAG GGC GGT C
	P2 T3+ R3B INSERT	CAG AGA TGC AAT TAA CCC TCA CTA AAG GGA GA AG AGA CTC ATG CCA CAG GCA AG
3.	Per2 T7+ F564	CCA AGC TTC TAA TAC GAC TCA CTA TAG GGA GA GAA CGC GGA TAT GTT TGC TG
	Per2 T3+R1146	CAG AGA TGC AAT TAA CCC TCA CTA AAG GGA GA TTT GTG GAT GGC GAG CAT CAA G
4.	P2 T7+ F1517	CCA AGC TTC TAA TAC GAC TCA CTA TAG GGA GA T CTA GCC GGT GGA GAT CCG A
	P2 T3+ R2216	CAG AGA TGC AAT TAA CCC TCA CTA AAG GGA GA CGG AAC TTC TGC AGG AAG C
5.	P2 T7+ F1821	CCA AGC TTC TAA TAC GAC TCA CTA TAG GGA GA CCT GAA GAG GAA GTG CGA GTT
	P2 T3+ R2216	CAG AGA TGC AAT TAA CCC TCA CTA AAG GGA GA CGG AAC TTC TGC AGG AAG C
6.	Vip T7+F202	CCA AGC TTC TAA TAC GAC TCA CTA TAG GGA GA GAC TCT CTT CAG TGT GCT STT CTC
	Vip T3+ R657	CAG AGA TGC AAT TAA CCC TCA CTA AAG GGA GA A AGT CGG GAG AAT CTC CCT CAC

RESULTS

1. Cloning and Characterization of *Per2* CDS from the SCN of *F. palmarum*

Primary PCR reaction using internal primers

To elucidate the complete coding region of *F.palmarum*, initially primers were designed from internal conserved domains of *Per2* transcript. Sense and Antisense primer pairs, F84/R1146 and F927/R2216 was used in separate PCR reactions with ZT 12 SCN cDNA to generate amplicons with an expected size of 1062 and 1289 bp respectively. The amplicons were excised and DNA was extracted to be used as template for 2nd round PCR. (Fig.17a, b). Further, pooling the amplicons had significantly increased the yield during gel extraction, the eluate with the specific primer was send for sequencing at CDFD, Hyderabad. The sequencing reaction confirmed the identity of the transcript as *Period2* (C5 and C6).

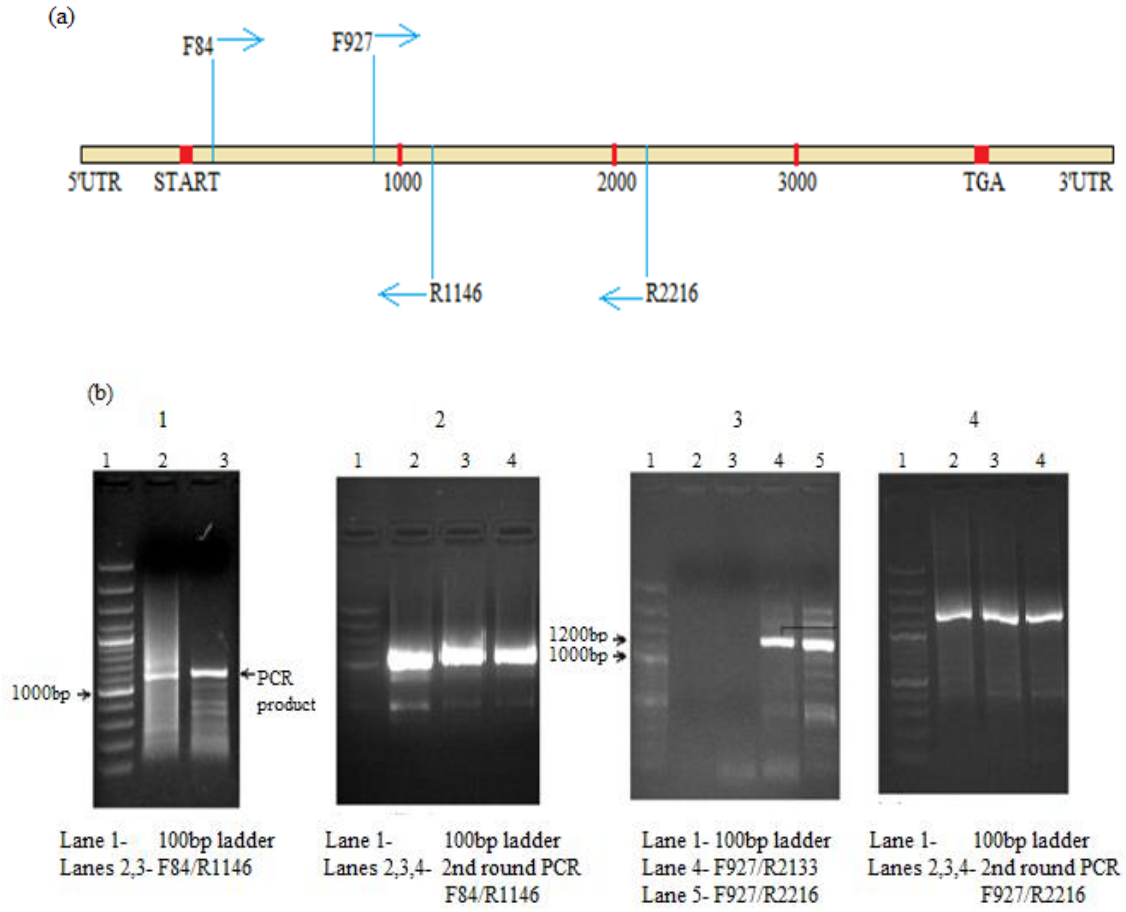


Fig. 17. (a) A diagrammatic representation pf Per2 mRNA indicating the sense and antisense primer region. (b) Agarose gel images showing PCR products (1) F84/R1146 (2) F927/R2216 (3) & (4) respective 2nd round PCR products.

gDNA extraction and PCR for elucidating the 5'end of the transcript

The conventional 5' and 3'RACE was not employed in the present work to know the 5' or 3' end of *F.palmarum Per2*. Instead, a comparitive approach based on the presence of 'conserved regulatory elements' in the untranslated region of the transcript

was explored (Fig. 19). A primer (F ConE) was hence designed based on an 'E'E box' reported to be present in the 5'UTR of rat, mouse, as well as primate *Per2*. This sense primer along with an antisense primer at Exon1 (R113) was predicted to give an 8-10kb amplicon with genomic DNA as it would span the 1st intron of the gene. An exonic architecture of *Per2* which was constructed for the study indicated relatively high conservation of exon and intron size and number across mouse, rat and humans (Fig.18). This was useful to predict expected size of PCR products when using Exon/Intron spanning primers.

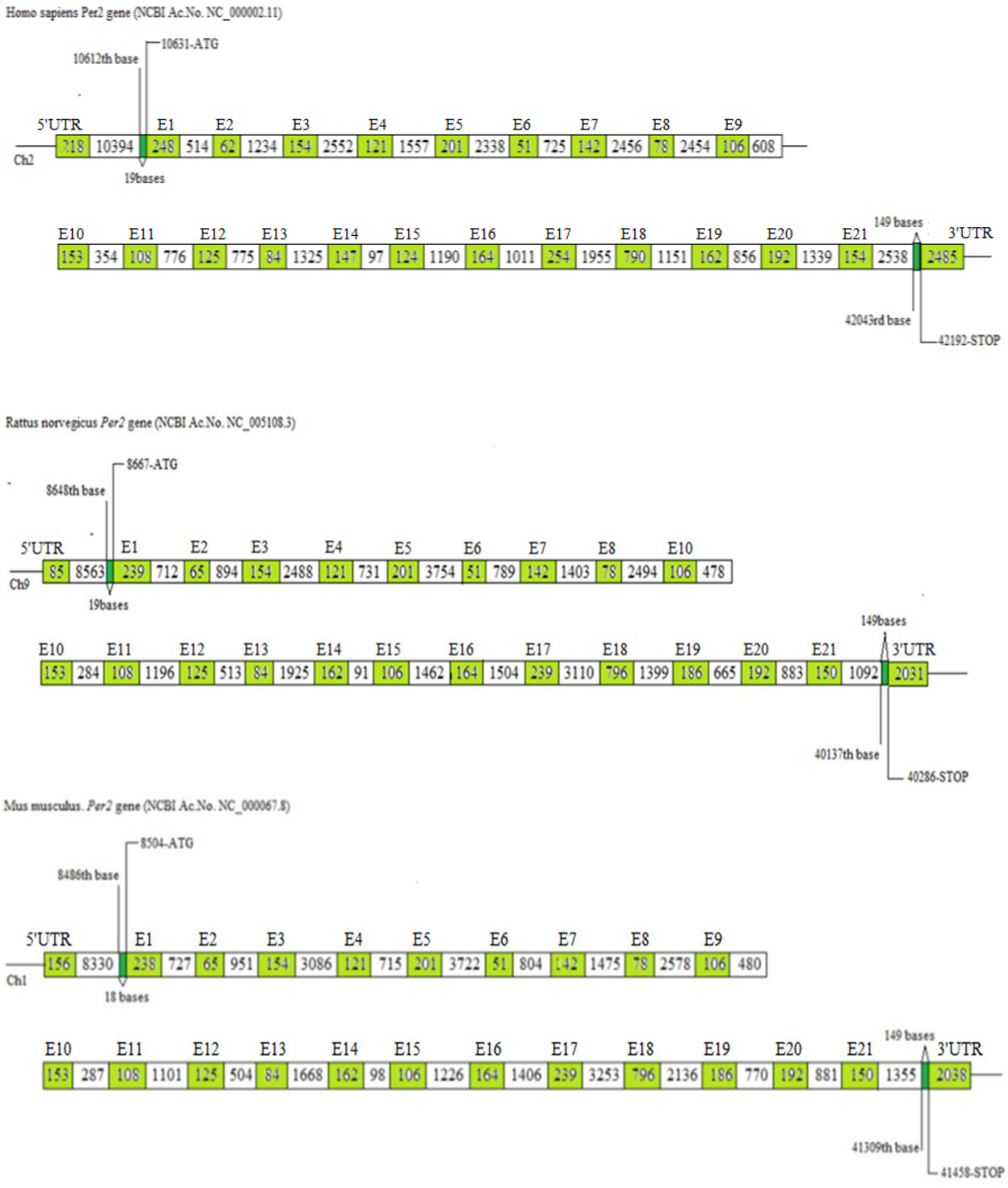


Fig.18. Exonic architecture of Per2 constructed for *Homo sapiens*, *Rattus norvegicus* and *Mus musculus*

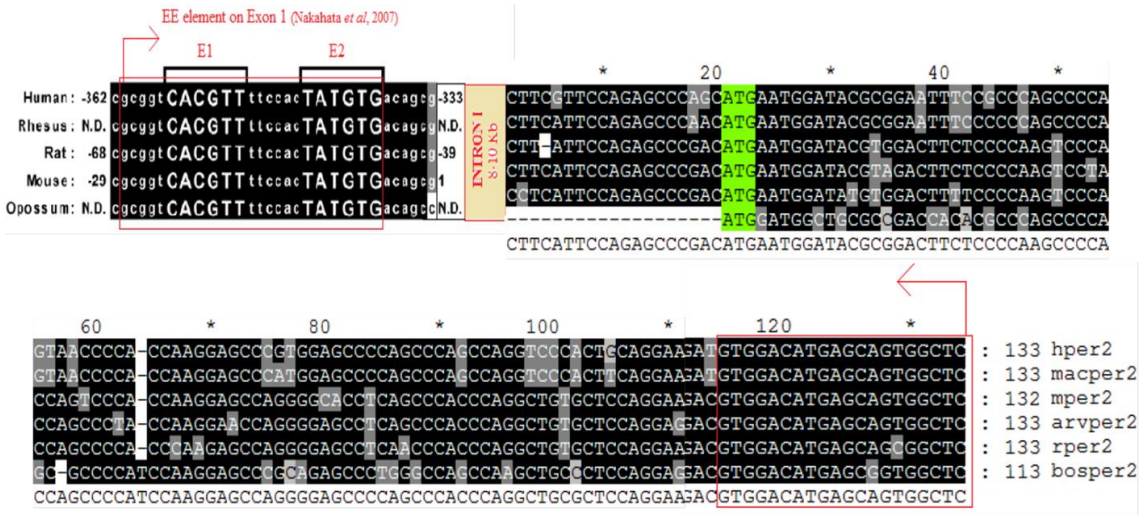


Fig. 19. Diagrammatic representation showing the E'E regulatory element and the antisense primer region.

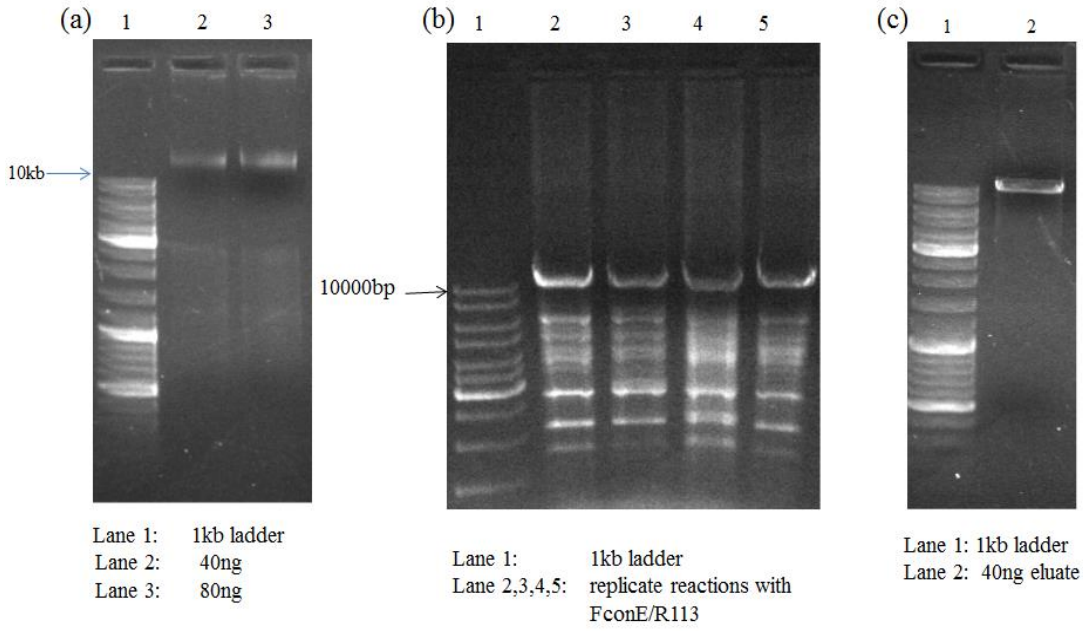


Fig.20. (a) gDNA extracted from the SCN of *F.palmarum*. (b)10 kb amplicon obtained with the primers F cone/R113 in Long Range PCR with gDNA as template.(c) Amplicons after Gel extraction loaded on agarose gel

The amplicons were excised (Fig 20.) pooled and the DNA extracted using commercial kit (Qiagen). The eluate was analysed by spectrophotometric reading (Nanodrop, Thermo scientific) and gel electrophoresis. 35ng of the same was given for sequencing. R113 primer was used as this would be the only exonic fragment available for comparison and identification of the transcript. It was also important as any genomic information on this rodent model is lacking.

BLAST analysis confirmed similarity in the 3'end of the PCR product to *Homo sapien Per2* (C1, B1). Moreover, the region indicated the presence of an alternate start

codon in place of ATG which was reproducible with repeated experiments (A1) and sequencing (Fig. 21).

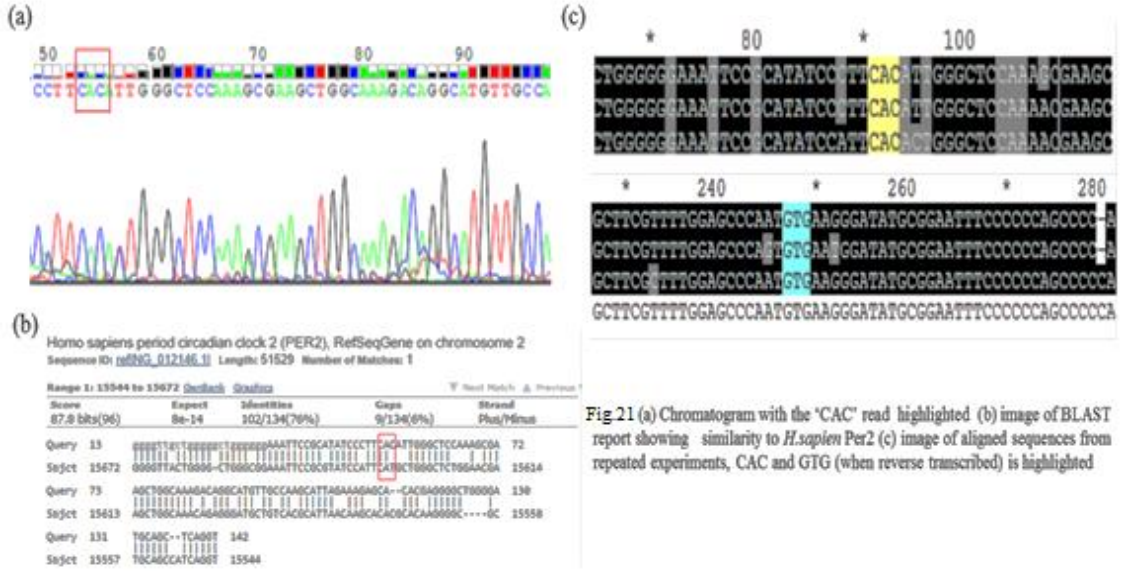


Fig 21 (a) Chromatogram with the 'CAC' read highlighted (b) image of BLAST report showing similarity to *H.sapien* Per2 (c) image of aligned sequences from repeated experiments, CAC and GTG (when reverse transcribed) is highlighted

FconE/R113 reaction was confirmed using well characterized rodent (c57BL6) and a wild rodent gDNA sample

Inorder to confirm the observation of an alternate 'Start' codon in *F.palmarum*, PCR reaction was repeated using SCN-gDNA isolated from c57BL6 mouse strain as well as wild mouse (Fig.22). The latter would assure that the results are not attributable to the use of a wild strain. C57BL6 genomic information being well established it would also confirm the specificity of the primers and hence the authenticity of the results obtained.

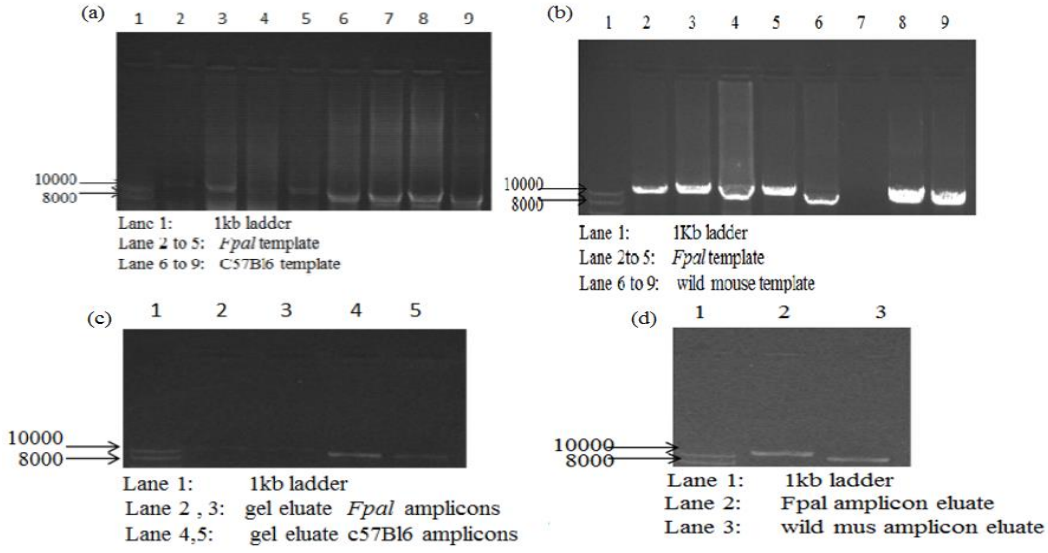


Fig. 22. amplicons from F cone/R113 PCR using c57bl6 (a) and wild mouse (b) SCN-gDNA template *F.palmarum* gDNA was used as positive control with both. (c) & (d) the gel extracted PCR products, from (a) and (b).

The sequencing data of the c57bl6 strain matched with the already existing information in NCBI, the start codon for *Per2* in this strain as known before is 'ATG'. The wild mouse *Per2* sequence was also very much comparable to the other laboratory rodent strains (C3, C4, D2, B2, B3, A2, A3) These reports confirmed the specificity of our primers to the region flanking the 1st intronic sequence of *Per2* (Fig. 23). Also it shows the extreme conservation and the possible importance of this region in clock function.

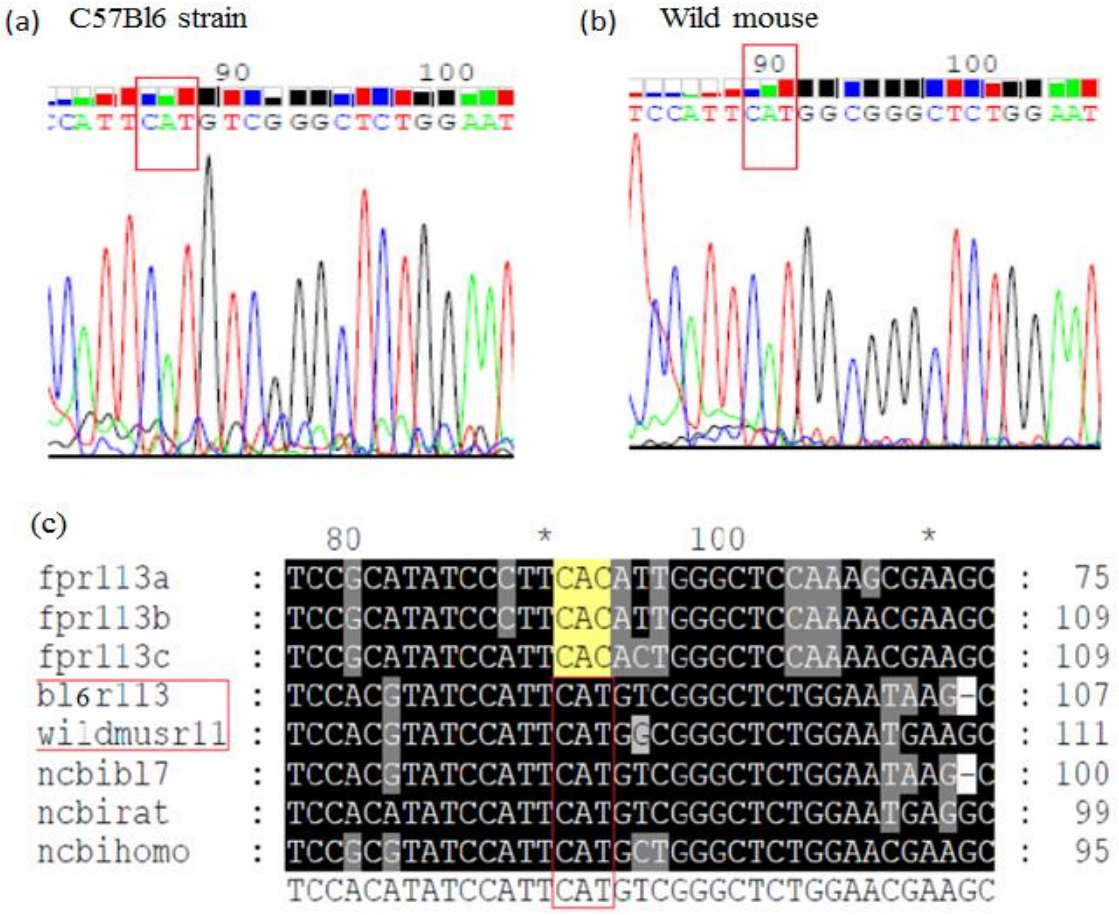


Fig.23. A segment of the chromatogram with the CAT (reverse transcribed as ATG) read highlighted in c56bl7 (a) and wild mouse (b). Short region of CLUSTAL alignment (c) with the lab generated sequencing data of c56bl7 (bl6 r113) and wild mouse (wildmusr113) compared with the *F.palmarum* sequence (fpr113,a,b,c) and NCBI data for the same region in mouse, rat and humans (ncbibl7, ncbirat, ncbihomo respectively).

TA cloning of a short segment comprising the 5'end of the *Per2* intergenic element and 113 bp downstream of the putative Start codon

A forward primer towards the 5'end of intron1 was used with the R113 primer to generate a short amplicon from the *F. palmarum* SCN-gDNA (Fig. 24). This was further cloned into a TA vector, the positive plasmids were isolated and further used for sequencing using both the vector specific and gene specific primers. This had improved the quality of chromatogram in comparison to using the gel eluates, also provided more reliable reading with the flexibility in using different primers (other than R113) for sequencing (C2). It was also confirmed that the start codon in *F.palmarum Per2* is 'GTG' and based on this data a new sense primer was designed [F Start1 (d)] just upstream to this 'Start codon' which along with an antisense would amplify the complete CDS of the transcript.

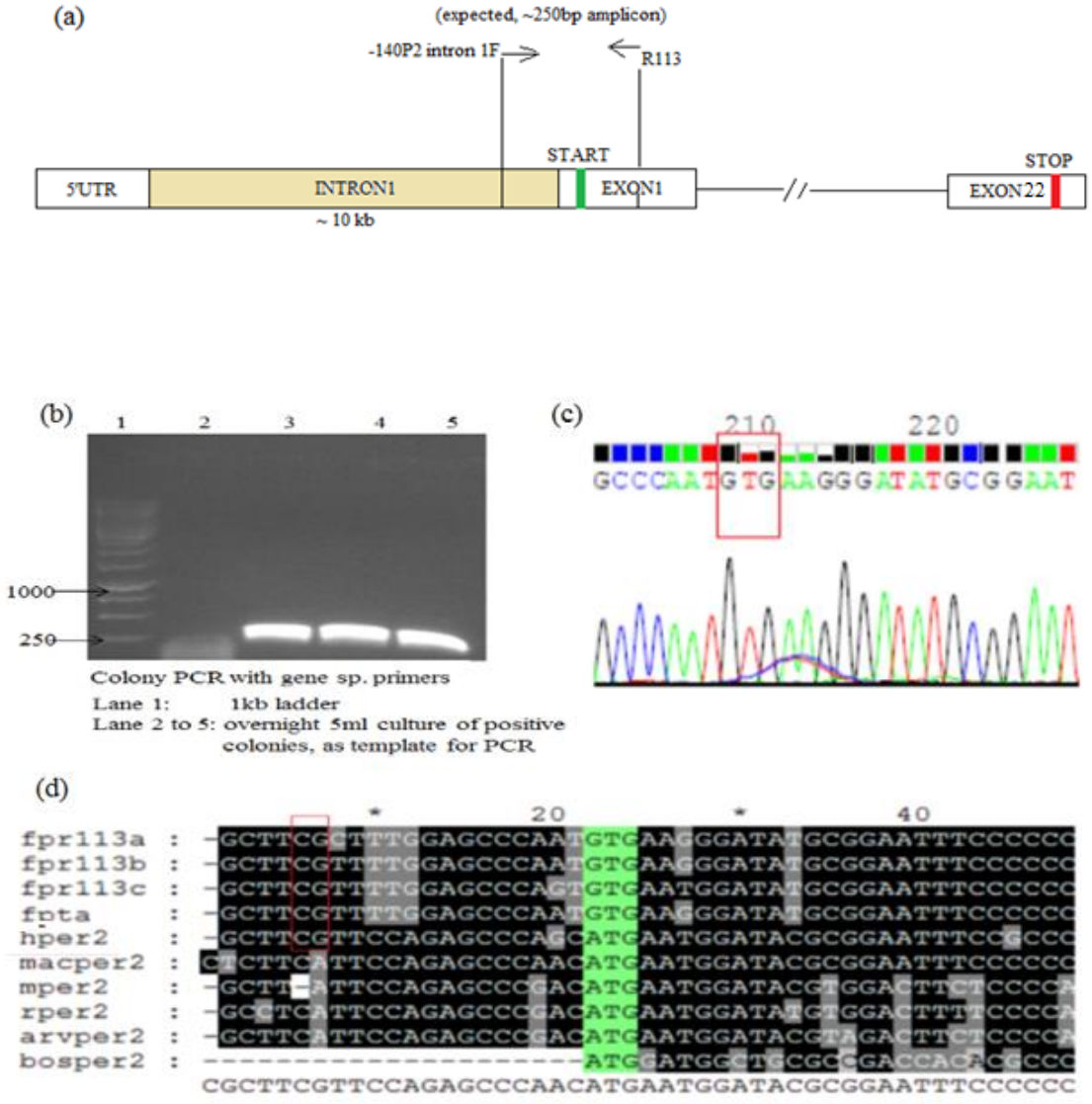


Fig.24. (a) Diagrammatic representation of *F.palmarum Per 2* gene showing the primer region (b) colony PCR with gene specific primers for selection of TA plasmid colonies with the insert. (c) small region of chromatogram (plasmid sequencing with vector specific primers) highlighting the 'GTG' at the

start codon position (d) alignment showing comparison of the region in *F.palmarum* [fp113 a,b and c, fpta (TA plasmid with the insert)] with *Homo sapiens* (hper2), *Macaca mullata* (macper2), *Bos taurus* (bosper2), *Arvicanthis* (arvper2), *Mus musculus* (musper2), *Rattus* (rper2). The intron-exon junction is 'boxed' in red.

Generating the complete coding information of *Per2* using an antisense primer based on a microRNA binding site at the 3'UTR of the transcript.

The role of miRNAs in regulating the mRNA rhythms of *Per2* is well investigated, miRNA binding to its 3'UTR being a general phenomenon CUSTAL W alignment showed the presence of many conserved nucleotide stretches within the region. An antisense primer was therefore designed at approximately 900 bp (R3'UTR 944) downstream to the putative stop codon (Fig. 25a). With the sense primer mentioned previously [F Start1(d)] this 3'UTR based reverse primer gave a specific and expected amplification between 4-5kb which was cloned and sequenced to confirm its identity (Fig. 25b).

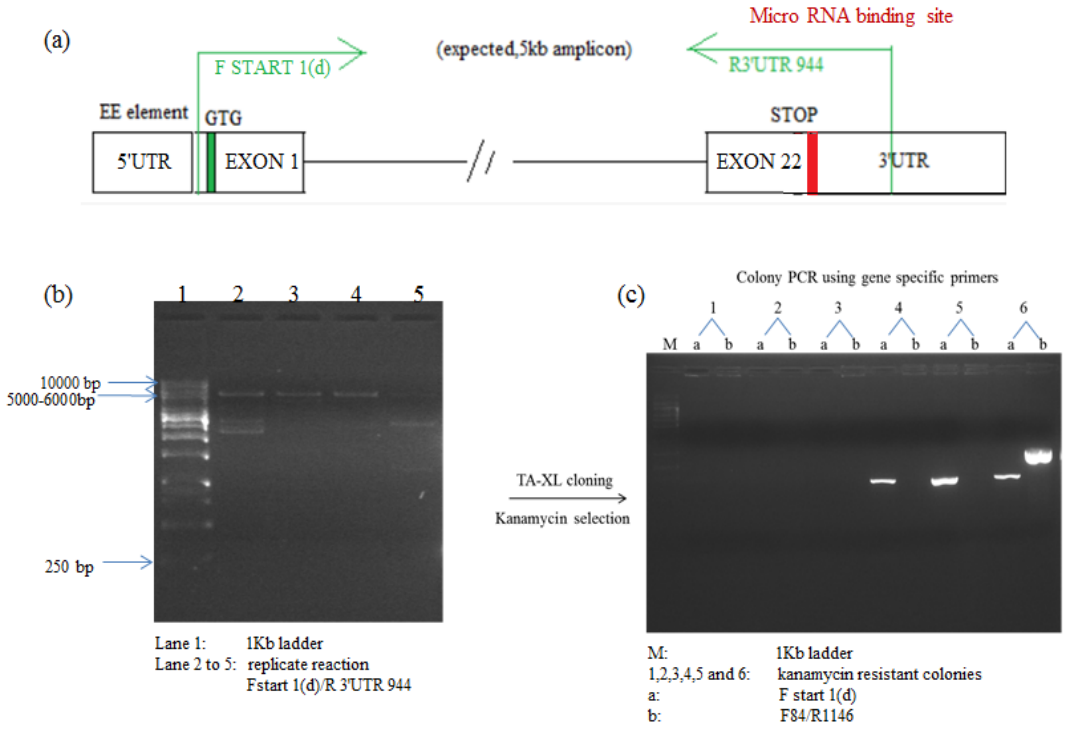


Fig.25. (a) pictorial representation of *Per2* mRNA with the primer region indicated. (b) Gel with the PCR products primed using F Start 1(d)/R3'UTR944 (c) Colony PCR for positive selection of TA-XL plasmids carrying the insert. Each colony was analysed using 2 primer pairs characterized previously, F start1 (d)/R610 and F84/R1146.

Plasmid was extracted from positive colonies using commercial kit (Qiagen), purity of the plasmid was confirmed by spectrophotometry. 150ng of the same was further used

for sequencing reaction. The work has employed a total of 21 primers (Fig. 26) to sequentially walk across the construct inorder to decipher the nucleotide sequence of the *Per2* CDS (C7 to C24).

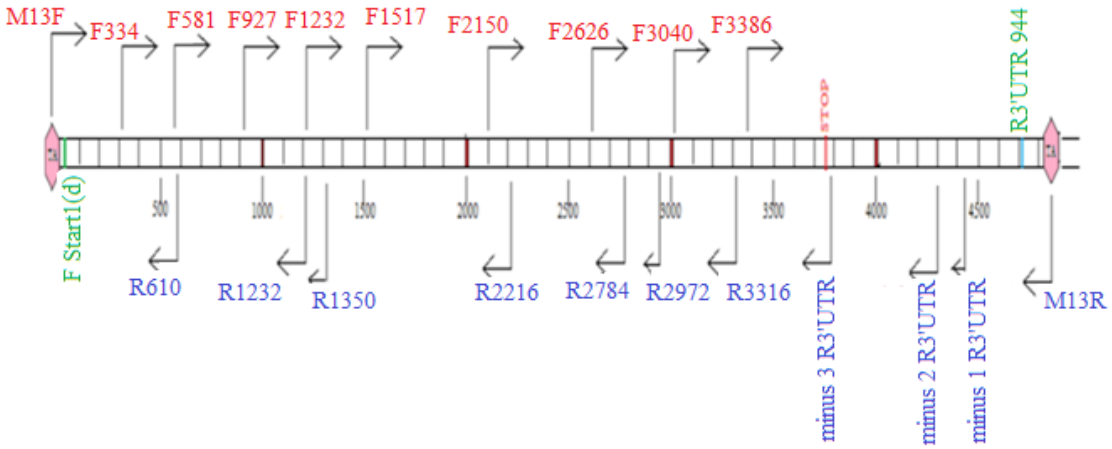


Fig.26. Diagram indicating the approximate position of individual primers across the *Per2* transcript which was used for gene sequencing.

Sequence Analysis

The putative Open Reading Frame (ORF) of *F. palmarum Per2* is of 3,756 bps. (D3, A4). The protein, with 'GTG' as the alternate start codon is composed of 1251 aminoacids. The nucleotide sequences shows 81% homology to both mouse and human transcripts. A conserved domain search (BLASTP, SMART, Motif-scan) within the translated sequence confirms that *F.pal* PER2 retains all the signature structural domains and functional motifs of the Period superfamily that are significant to dictate the molecular purpose of the protein. The PAS (Per-Arnt-Sim) domain comprising of the PAS-A, PAS-B and the PAS-like PAC motifs were formed within the 180- 450 amino acid stretch of the protein. The algorithms could also detect 'heme binding pockets' dispersed within the PAS domains. The PAS-A domain showed 81% and 82% homology with that of muridae and hPER2, respectively. PAS-B and PAC were more than 90% homologous to both murine and human PER2. Bipartite NLS , conserved phosphorylation motifs, proline (aminoacisd 839- 960) and serine rich regions, c-terminal Period_C domain etc. were also predicted using these tools. Importantly, the FASP motif is formed between 655-674th amino acid residues, the structure of this region is very well conserved in this animal model as well.

Transcript Variant of *Per2* (represented as *F. pal Per2v)*

A variant of the normal *Per2* was detected which was only 4116bp in length, mainly due to an insertion of 503 nucleotides followed downstream by a deletion of 1097 nucleotides (C25, C26) The region and nature of insertion and deletion were predicted based on *Per2* genomic DNA data of Mus, Rattus and Homo sapiens. The insertion which is between the exons 9 and 10, could therefore be an event of Intronic Retention (IR) which introduces a PTC (Premature Termination Codon) in the transcript, however the translated sequence retains the PAS domain and the truncated protein will have only 454 amino acid residues. The exons 14, 15 16 and 17 has been deleted in this transcript (A5, D4).

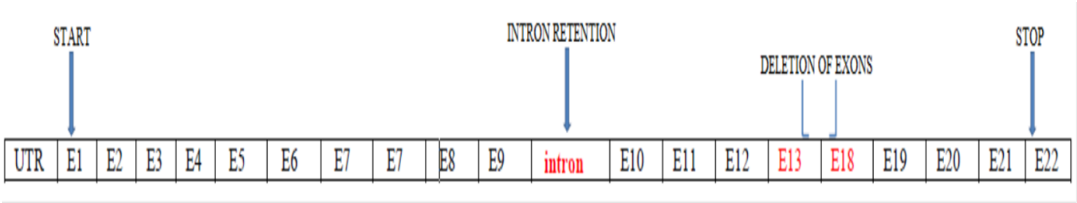
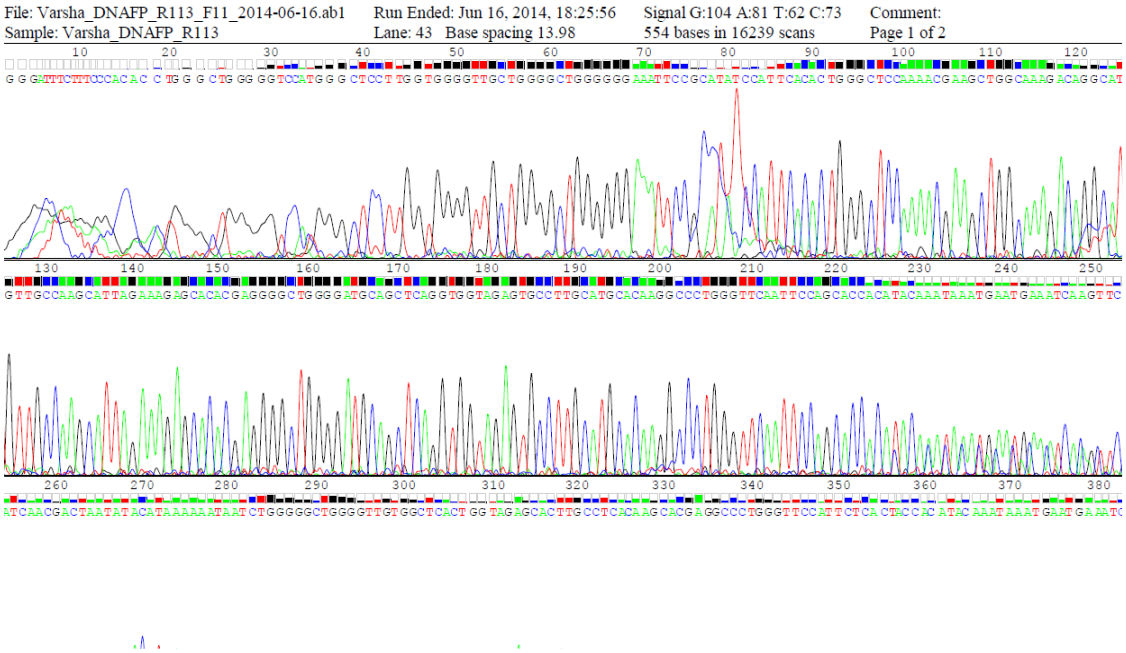
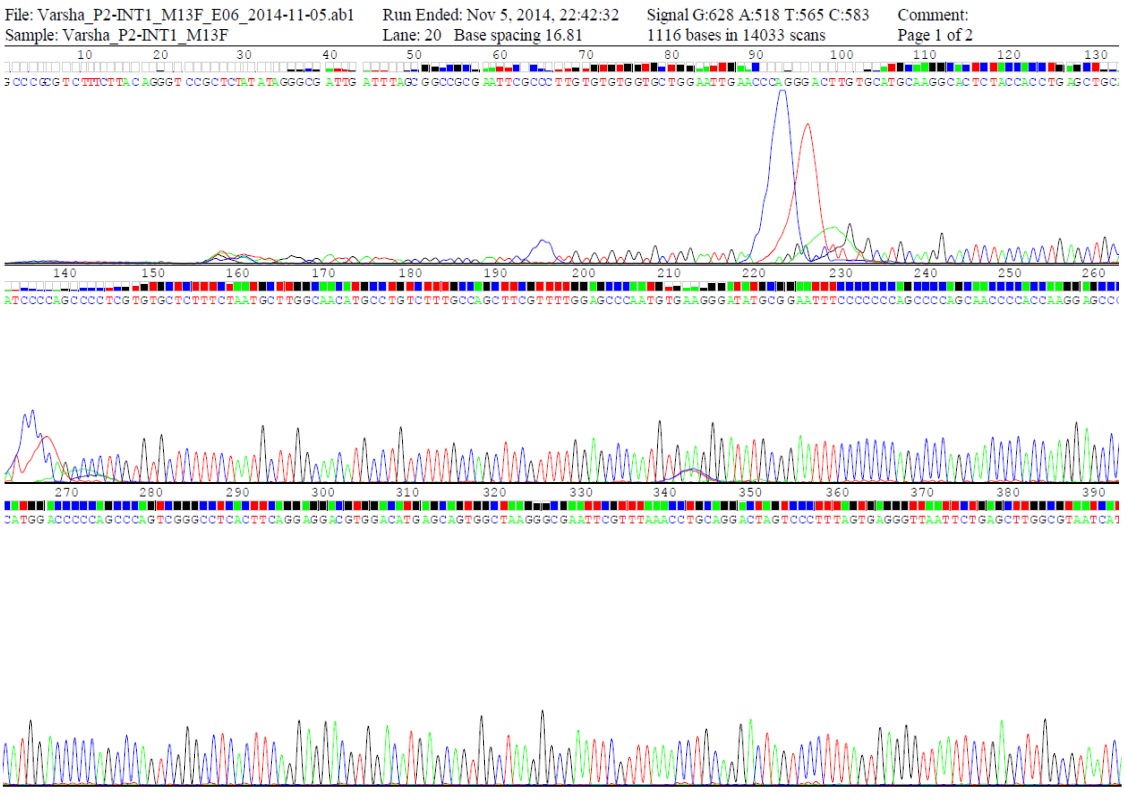


Fig.27. Diagrammatic representation of the *F.palmarum Per2* variant (*F.pal Per2v*)

C1. gDNA *F.palmarum* FconE/R113 PCR eluate sequencing. Primer R113



C2. TA clone with -140P2 intron 1F- R113 fragment, sequencing with Primer M13F



D1. Nucleotide sequences exported from the chromatogram C1 and C2

The CAC/GTG is marked in the sequences, underlined are the upstream and downstream exon sequences

>Varsha_DNAFP_R113 sequence exported from Varsha_DNAFP_R113_F11_2014-06-16.ab1

GGGATTTCTTTCCCACACCTGGGCTGGGGGTCCATGGGCTCCTTGGTGGGGT
TGCTGGGGCTGGGGGGAAATTCCGCATATCCATTCACACTGGGCTCCAAA
ACGAAGCTGGCAAAGACAGGCATGTTGCCAAGCATTAGAAAGAGCACACG
AGGGGCTGGGGATGCAGCTCAGGTGGTAGAGTGCCTTGCATGCACAAGGCC
CTGGGTTCAATTCCAGCACCACATACAAATAAATGAATGAAATCAAGTTCA
TCAACGACTAATATACATAAAAAATAATCTGGGGGCTGGGGTTGTGGCTCA
CTGGTAGAGCACTTGCCTCACAAGCACGAGGCCCTGGGTTCCATTCTCACT
ACCACATACAAATAAATGAATGAAATCAAGTTCTCAACAACTAATATATAT
AAAAAATAAGCTGGGCATGGCGGCAACTGCCTGTACCCCAGCTGTGGAGG
GAGCTAAATCAGAAGGATGATAAGCCCCACCCCACCTCATCAAAAAAGTG
AGGCGCTACCTGCTTAATGAGAGCCAGGCCTCTAAAAAAAAAAATTT

Reverse Complement ('GTG' is highlighted)

AAATTTTTTTTTTTAGAGGCCTGGCTCTCATTAAGCAGGTAGCGCCTCACTT
TTTTGATGAGGTGGGGTGGGGCTTATCATCCTTCTGATTTAGCTCCCTCCAC
AGCTGGGGTACAGGCAGTTGCCGCCATGCCCAGCTTATTTTTTATATATATT
AGTTGTTGAGAACTTGATTTCATTCATTTATTTGTATGTGGTAGTGAGAATG
GAACCCAGGGCCTCGTGCTTGTGAGGCAAGTGCTCTACCAGTGAGCCACAA
CCCCAGCCCCCAGATTATTTTTTATGTATATTAGTCGTTGATGAACTTGATTT
CATTCATTTATTTGTATGTGGTGCTGGAATTGAACCCAGGGCCTTGTGCATG
CAAGGCACTCTACCACCTGAGCTGCATCCCCAGCCCCTCGTGTGCTCTTTCT
AATGCTTGGCAACATGCCTGTCTTTGCCAGCTTCGTTTTGGAGCCCAGTGT
GAATGGATATGCGGAATTTCCCCCCAGCCCCAGCAACCCCACCAAGGAGC
CCATGGACCCCCAGCCCAGGTGTGGGAAAGAAATCCC

Nucleotide sequences exported from the chromatogram C2

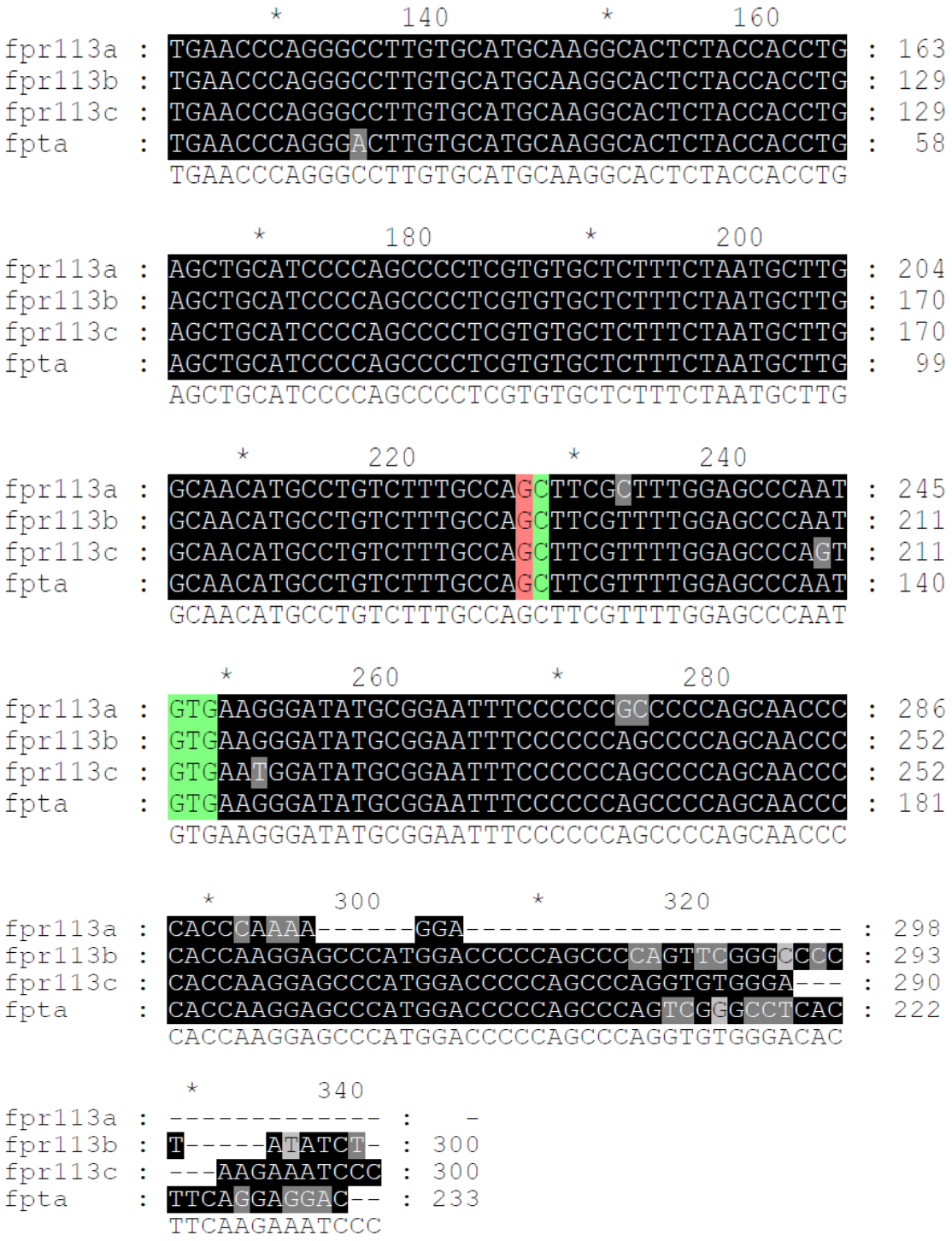
The primer -140P2 intron 1F and R113 is highlighted.

GCCCGCGTCTTTCTTACAGGGTCCGCTCTATATAGGGCGATTGATTTAGCGG
CCGCGAATTCGCCCTTGTGTGTGGTGCTGGAATTGAACCCAGGGACTTGTG
CATGCAAGGCACTCTACCACCTGAGCTGCATCCCCAGCCCCTCGTGTGCTCT
TTCTAATGCTTGGCAACATGCCTGTCTTTGCCAGCTTCGTTTTGGAGCCCAA

TGTGAAGGGATATGCGGAATTTCCCCCCCAGCCCCAGCAACCCCACCAAG
GAGCCCATGGACCCCCAGCCCAGTCGGGCCTCACTTCAGGAGGACGTGGAC
ATGAGCAGTGGCTAAGGGCGAATTCGTTTAAACCTGCAGGACTAGTCCCTT
TAGTGAGGGTTAATTCTGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTG
TGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCA
TAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG
CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCA
TTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTC
TTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGA
GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGG
GATAACGCACGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA
CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGAC
GAGCATCACAAAAATCGACGCTCAAGTCAGAGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCTGT
CCGACCCTGCCGCTTACCGGAAACCTGTCCGCCTTTCTCCCTTCGGGAAGCG
TGGCGCTTTCTCAATAGCTCACGCTGTAGGTATCTCAGTCGGGTGTAAGTCG
TTCGCTCCAAGCTGGGCTGGTGGTGCACGAACCCCCGTCAGCGACGCTGGC
GCCTTATCGGTAACTATCGCTTGGAGTCCAACCGGTAG

A1. Alignment of 3 sequencing data (a,b,c) of *F.palmarum* gDNA FconE/R113 PCR gel eluates with TA clone with -140P2 intron 1F- R113 fragment to generate consensus sequence

```
                  *         20         *         40
fpr113a  :  -CCCACTGTATAAAGTTCCAAGAAAATGTGTAGAAAGATGC  :  40
fpr113b  :  CCCTTTCTTGCGGGGTCACCCGG------------------  :  23
fpr113c  :  CACAACCCCA----GCCCCCAGA------------------  :  19
fpta     :  ----------------------------------------  :   -
            CCCAACCTTACAAAGTCCCCAGAAAATGTGTAGAAAGATGC

                  *         60         *         80
fpr113a  :  GAAAGGGTGGGCACCCCCACACGGCCGGTGGTCTCCCTCAC  :  81
fpr113b  :  ----------GCACCCTCGCGTGTGTGTGTGT-------GT  :  47
fpr113c  :  ----------TTATTTTT-TATGTATATTAGTCGT--TGAT  :  47
fpta     :  ----------------------------------------  :   -
            GAAAGGGTGGGCACCCTCACATGTATGTTAGTCGCCCTCAT

                  *        100         *        120
fpr113a  :  ATGCGCGTGTGTGTGTGTGTGTGTGTGTGTGGTGCTGGAAT  : 122
fpr113b  :  GTGTGTGTGTGTGTGTGTGTGTGTGTGTGGTGCTGGAAT  :  88
fpr113c  :  GAACTTGATTTCATTCATTTATTTGTATGTGGTGCTGGAAT  :  88
fpta     :  ----------------------GTGTGTGGTGCTGGAAT  :  17
            GTGCGTGTGTGTGTGTGTGTGTGTGTGTGTGGTGCTGGAAT
```



Consensus sequence. fpr113a,b,c with TAclone

CCCAACCTTACAAAGTCCCCAGAAAATGTGTAGAAAGATGCGAAAGGGTG
GGCACCCTCACATGTATGTTAGTCGCCCTCATGTGCGTGTGTGTGTGTGTGT
GTGTGTGTGTGGTGCTGGAATTGAACCCAGGGCCTTGTGCATGCAAGGCAC
TCTACCACCTGAGCTGCATCCCCAGCCCCTCGTGTGCTCTTTCTAATGCTTG
GCAACATGCCTGTCTTTGCCAGCTTCGTTTTGGAGCCCAATGTGAAGGGAT
ATGCGGAATTTCCCCCCAGCCCCAGCAACCCCACCAAGGAGCCCATGGACC
CCCAGCCCAGGTGTGGGACACTTCAAGAAATCCC

B1. NCBI-BLAST report of consensus sequence- hits to *Per2* is enlisted

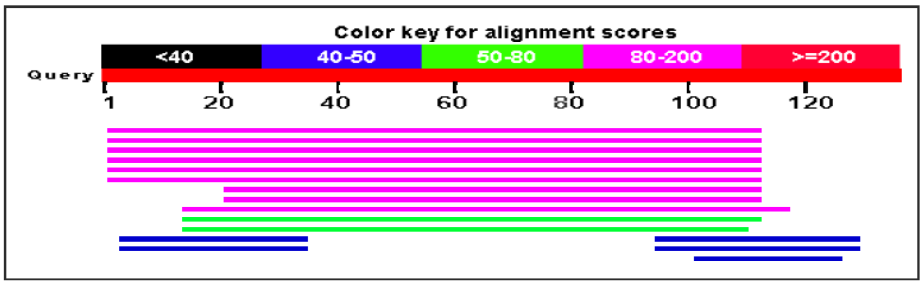
Nucleotide Sequence (136 letters)

RID PRGNUTR501R (Expires on 06-24 19:06 pm)
Query ID lcl|Query_21399
Description None
Molecule type nucleic acid
Query Length 136

Database Name nr
Description Nucleotide collection (nt)
Program BLASTN 2.4.0+

Graphic Summary

Distribution of 16 Blast Hits on the Query Sequence



Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: Papio anubis period circadian clock 2 (PER2), transcript variant X4, mRNA	138	138	81%	3e-29	87%	XM_009183466.1
PREDICTED: Aotus nancymaae period circadian clock 2 (PER2), transcript variant X1, mRNA	129	129	81%	1e-26	86%	XM_012473480.1
Homo sapiens period circadian clock 2 (PER2), RefSeqGene on chromosome 2	129	129	81%	1e-26	86%	NG_012146.1
Homo sapiens period 2 (PER2) gene, complete cds	129	129	81%	1e-26	86%	EF015905.1
Homo sapiens period circadian protein 2 (PER2) gene, partial cds	129	129	81%	1e-26	86%	AY647991.1
Homo sapiens BAC clone RP11-5O24 from 2, complete sequence	129	129	81%	1e-26	86%	AC012485.13
PREDICTED: Marmota marmota marmota period circadian protein homolog 2-like (LOC107138849), mRNA	120	120	66%	7e-24	89%	XM_015481656.1
PREDICTED: Ictidomys tridecemlineatus period circadian clock 2 (Per2), mRNA	111	111	66%	4e-21	87%	XM_013359167.1
PREDICTED: Miniopterus natalensis period circadian clock 2 (PER2), mRNA	87.8	87.8	75%	4e-14	80%	XM_016203632.1
PREDICTED: Equus przewalskii period circadian clock 2 (PER2), mRNA	75.2	75.2	72%	3e-10	78%	XM_008533309.1
PREDICTED: Ursus maritimus period circadian clock 2 (PER2), mRNA	71.6	71.6	70%	3e-09	77%	XM_008690278.1

An alignment of the region with *hper2*

Homo sapiens period circadian protein 2 (PER2) gene, partial cds
Sequence ID: **gb|AY647991.1|** Length: 536 Number of Matches: 1
Range 1: 303 to 413

Score	Expect	Identities	Gaps	Strand	Frame
129 bits(142)	1e-26()	95/111(86%)	0/111(0%)	Plus/Plus	

Features:

```
Query  2    CAACATGCCTGTCTTTGCCAGCTTCGTTTTGGAGCCCAATGTGAAGGGATATGCGGAATT  61
            ||| ||| ||| | |||||||||||||||||    ||||||||   |||| ||||| |||||||||
Sbjct  303  CAGCATCCCTCTGTTTGCCAGCTTCGTTCCAGAGCCCAGCATGAATGGATACGCGGAATT  362

Query  62   Tcccccccagccccagcaaccccaccaaggagcccatggaccccccagccccag  112
            ||| |||||||||||| ||||||||||||||||||| |||| |||||||||||
Sbjct  363  TCCGCCCAGCCCCAGTAACCCCACCAAGGAGCCCGTGGAGCCCCAGCCCAG  413
```

GenBank data indicating the beginning of CDS in *hper2* which is aligned with the query sequence

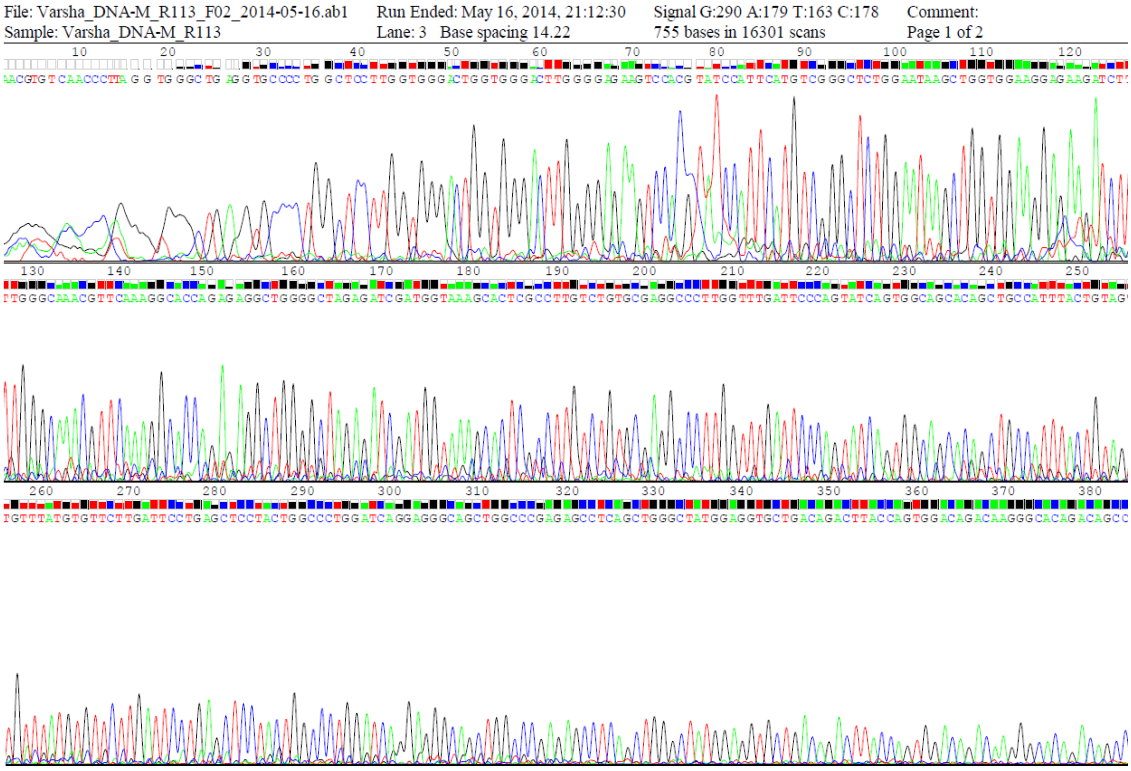
Homo sapiens period circadian protein 2 (PER2) gene, partial cds

GenBank: AY647991.1

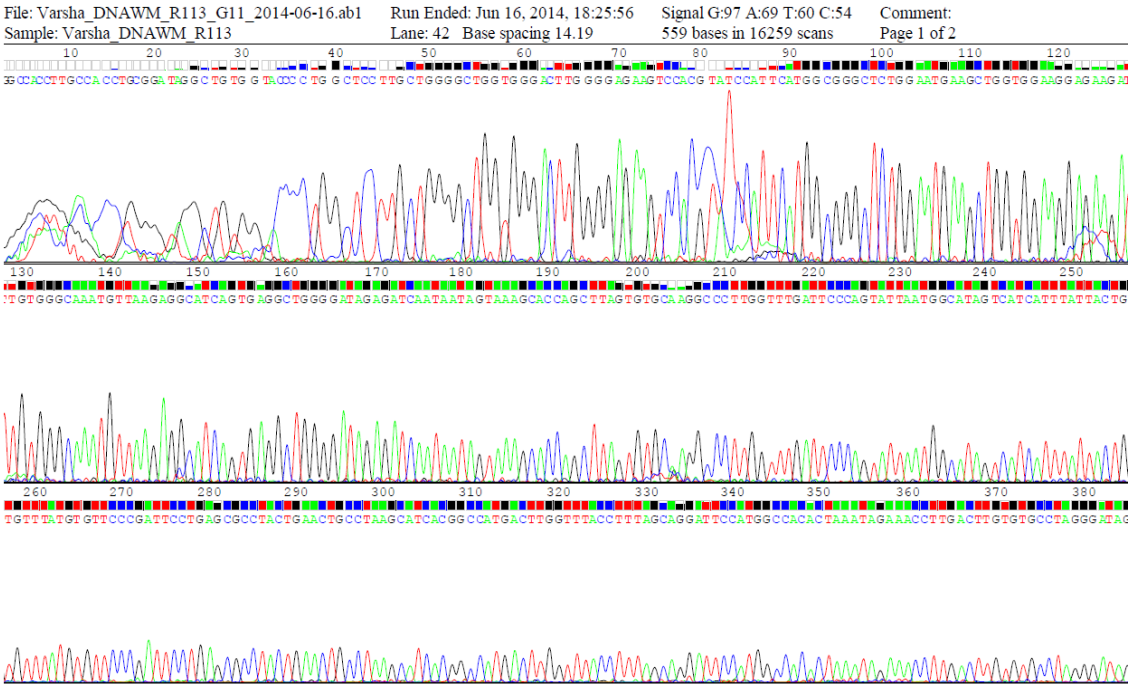
FASTA Graphics

```
     gene            <1..>536
                     /gene="PER2"
     variation       186..193
                     /gene="PER2"
                     /note="8-basepair repeat insertion"
                     /replace=""
     mRNA            <323..>536
                     /gene="PER2"
                     /product="period circadian protein 2"
     exon            323..>536
                     /gene="PER2"
                     /number=2
     CDS             343..>536
                     /gene="PER2"
                     /codon_start=1
                     /product="period circadian protein 2"
                     /protein_id="AAT68170.1"
                     /db_xref="GI:49619171"
                     /translation="MNGYAEFPPSPSNPTKEPVEPQPSQVPLQEDVDMSSGSSGHETN
                     ENCSTGRDSQGSDCDDSGKEL"
ORIGIN
        1 ggcaaaaccc tgtctctacc aaaaatacaa aaaattaacc aggcatggtg gtgagtgcct
       61 gtggttccag ctaccctgga ggctgaggtg ggaggatcgc ttgagcccag gcactgcatt
      121 ccctgggcaa cagagccaga ccctgtctcc aaagtagaaa gaaagaaaga gagagagaga
      181 gagacagaga gacagagaga agaaagaaag agactgcagt ttcattagta acagaaagag
      241 tcaaatgggt gctctacctg atggctgcag cgccccttgt gcgtgtgctt gttaatgcgt
      301 gacagcatcc ctctgtttgc cagcttcgtt ccagagccca gcatgaatgg atacgcggaa
      361 tttccgccca gccccagtaa ccccaccaag gagcccgtgg agccccagcc cagccaggtc
      421 ccactgcagg aagatgtgga catgagcagt ggctccagtg gacatgagac caacgaaaac
      481 tgctccacgg ggcgggactc gcagggcagt gactgtgacg acagtgggaa ggagct
//
```

C3. gDNA C57BL/6 FconE/R113 PCR eluate sequencing. Primer R113



C4. gDNA wild mouse FconE/R113 PCR eluate sequencing. Primer R113



D2. Nucleotide sequences exported from the chromatograms C3 and C4.

>Varsha_DNA-M_R113 sequence exported from Varsha_DNA-M_R113_F02_2014-05-16.ab1

The CAT/ATG (when reverse transcribed) is marked in the sequences, underlined are the upstream and downstream exon sequences

AACGTGTCAACCCTTAGGTGGGCTGAGGTGCCCCTGGCTCCTTGGTGGGAC
TGGTGGGACTTGGGGAGAAGTCCACGTATCCATT**CAT**GTCGGGCTCTGGA
ATAAGCTGGTGGAAGGAGAAGATCTTTGGGCAAACGTTCAAAGGCACCAG
AGAGGCTGGGGCTAGAGATCGATGGTAAAGCACTCGCCTTGTCTGTGCGAG
GCCCTTGGTTTGATTCCCAGTATCAGTGGCAGCACAGCTGCCATTTACTGTA
GTGTTTATGTGTTCTTGATTCCTGAGCTCCTACTGGCCCTGGATCAGGAGGG
CAGCTGGCCCGAGAGCCTCAGCTGGGCTATGGAGGTGCTGACAGACTTACC
AGTGGACAGACAAGGGCACAGACAGCCACCCTACATGACAGTGGGGAAAA
TATAGCCATGCTAAATGGTAAACCAGGCATGAGAATCTTCATCCTATTATTT
CCACAGGTGACATTAAGCAGTGGGGTCCAAAAACAACTCCATCACAACATG
GCATGTGCTACAGAGAAGCTGGGGGCAGCCGGGTCTCCTGGAGAGGTAAT
GCTGTTGGGTGAGGGGTGAAGAGGCCTCTATGCCATGGAGCGCTCCAGGTG
CTAGGGGCTCCGGCGCTGCAGGGAGCCTGGCATTGGAATAAGATGGTGTAA
GCGGGAAGGAGACTCAAGCTTACCCTGAGTTCGTCAGGGGCCCCTGAAGTG
CCTAAGCATCAAATCAGGGCGTGGTTTACATTGGGTCGGATC

>Varsha_DNAWM_R113 sequence exported from Varsha_DNAWM_R113_G11_2014-06-16.ab1

GGCCACCTTGCCACCTGCGGATAGGCTGTGGTACCCCTGGCTCCTTGCTGGG
GCTGGTGGGACTTGGGGAGAAGTCCACGTATCCATT**CAT**GGCGGGCTCTG
GAATGAAGCTGGTGGAAGGAGAAGATTGTGGGCAAATGTTAAGAGGCATC
AGTGAGGCTGGGGATAGAGATCAATAATAGTAAAGCACCAGCTTAGTGTGC
AAGGCCCTTGGTTTGATTCCCAGTATTAATGGCATAGTCATCATTTATTACT
GTGTTTATGTGTTCCCGATTCCTGAGCGCCTACTGAACTGCCTAAGCATCAC
GGCCATGACTTGGTTTACCTTTAGCAGGATTCCATGGCCACACTAAATAGA
AACCTTGACTTGTGTGCCTAGGGATAGGTGGGCAGCAGCTAAAGCCACGCT
GCTAAGAAGGCTTCCATAGACATCTGAAGGCCATGAAGGGTCACAGACTTA
GTGATTGGGAAGGATGGACATATCACCAGACAGGAGGTCACAGCTCAAGG
ATAAGAGACTGTTTCATGTCACTGGGAGAAAAAGGGAGGGGTCAGGGAT

B2. NCBI-BLAST report of C3 - specific hits to *mPer2* is enlisted

Nucleotide Sequence (112 letters)

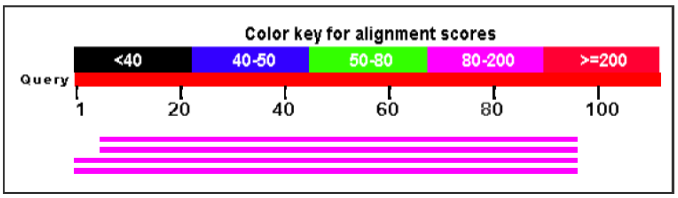
RID PR4BYB0G014 (Expires on 06-24 15:36 pm)

Query ID lcl|Query_1695
Description None
Molecule type nucleic acid
Query Length 112

Database Name Mouse G+T (2 databases)
Description
Program BLASTN 2.4.0+

Graphic Summary

Distribution of 4 Blast Hits on the Query Sequence



Description	Max score	Total score	Query cover	E value	Ident	Accession
Transcripts						
PREDICTED: Mus musculus period circadian clock 2 (Per2), transcript variant X1, mRNA	169	169	81%	6e-40	100%	XM_006529249.2
Mus musculus period circadian clock 2 (Per2), mRNA	169	169	81%	6e-40	100%	NM_011066.3

Alignment of the query sequence with *mPer2*, the ATG is highlighted

PREDICTED: Mus musculus period circadian clock 2 (Per2), transcript variant X1, mRNA
Sequence ID: **ref|XM_006529249.2|** Length: 5902 Number of Matches: 1
Range 1: 221 to 311

Score	Expect	Identities	Gaps	Strand	Frame
169 bits(91)	6e-40()	91/91(100%)	0/91(0%)	Plus/Plus	

Features:

```
Query  6    GCTTATTCCAGAGCCCGACATGAATGGATACGTGGACTTCTCCCCAAGTCCCACCAGTCC  65
Sbjct  221  GCTTATTCCAGAGCCCGACATGAATGGATACGTGGACTTCTCCCCAAGTCCCACCAGTCC  280
Query  66   CACCAAGGAGCCAGGGGCACCTCAGCCCACC  96
Sbjct  281  CACCAAGGAGCCAGGGGCACCTCAGCCCACC  311
```

B3. NCBI-BLAST report of C4 - specific hits to *mPer2* is enlisted

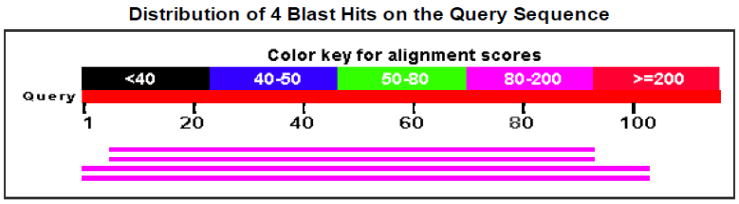
Nucleotide Sequence (116 letters)

RID PR4EH916014 (Expires on 06-24 15:37 pm)

Query ID lcl|Query_40963
Description None
Molecule type nucleic acid
Query Length 116

Database Name Mouse G+T (2 databases)
Description
Program BLASTN 2.4.0+

Graphic Summary



Description	Max score	Total score	Query cover	E value	Ident	Accession
Transcripts						
PREDICTED: Mus musculus period circadian clock 2 (Per2), transcript variant X1, mRNA	128	128	75%	1e-27	93%	XM_006529249.2
Mus musculus period circadian clock 2 (Per2), mRNA	128	128	75%	1e-27	93%	NM_011066.3

Alignment of the query sequence with *mPer2*, the ATG is highlighted

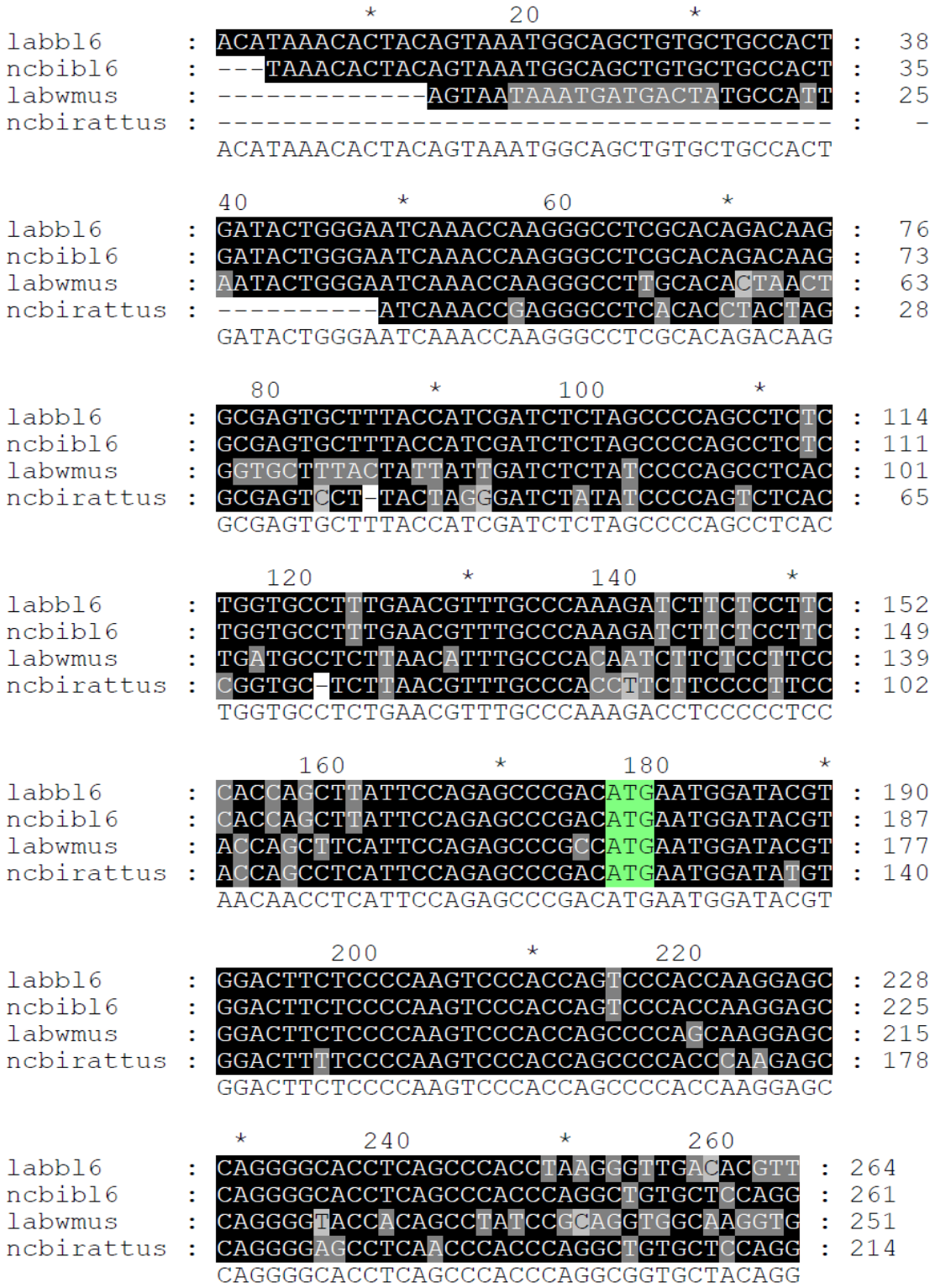
PREDICTED: Mus musculus period circadian clock 2 (Per2), transcript variant X1, mRNA
Sequence ID: **ref|XM_006529249.2|** Length: 5902 Number of Matches: 1
Range 1: 221 to 307

Score	Expect	Identities	Gaps	Strand	Frame
128 bits(69)	1e-27()	82/88(93%)	1/88(1%)	Plus/Plus	

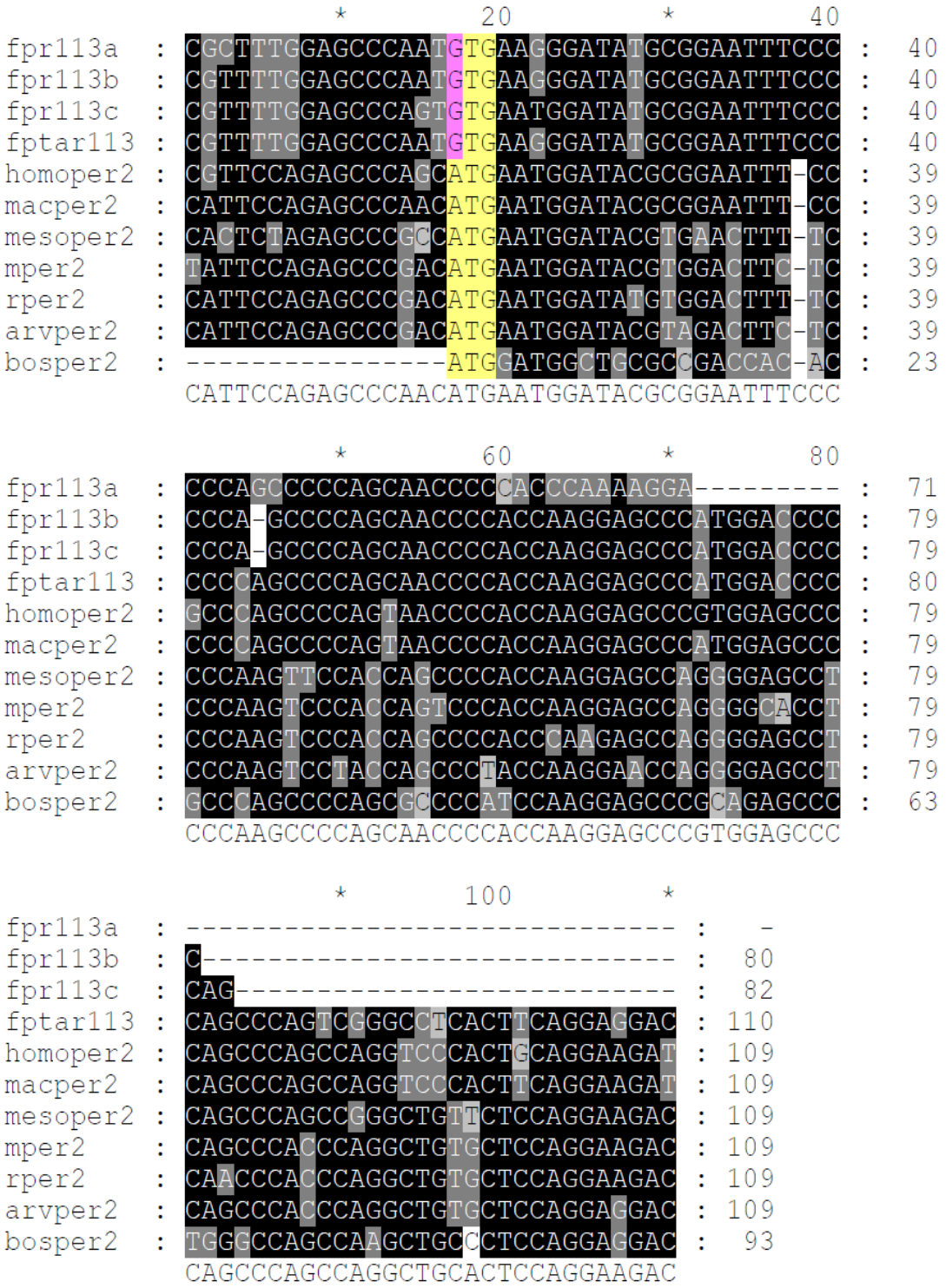
Features:

```
Query  6    GCTTCATTCCAGAGCCCGCCATGAATGGATACGTGGACTTCTCCCCAAGTCCCACCAGCC  65
            ||||  |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct  221  GCTT-ATTCCAGAGCCCGACATGAATGGATACGTGGACTTCTCCCCAAGTCCCACCAGTC  279

Query  66   CCAGCAAGGAGCCAGGGGTACCACAGCC  93
            ||| |||||||||||||| ||| ||||||
Sbjct  280  CCACCAAGGAGCCAGGGGCACCTCAGCC  307
```

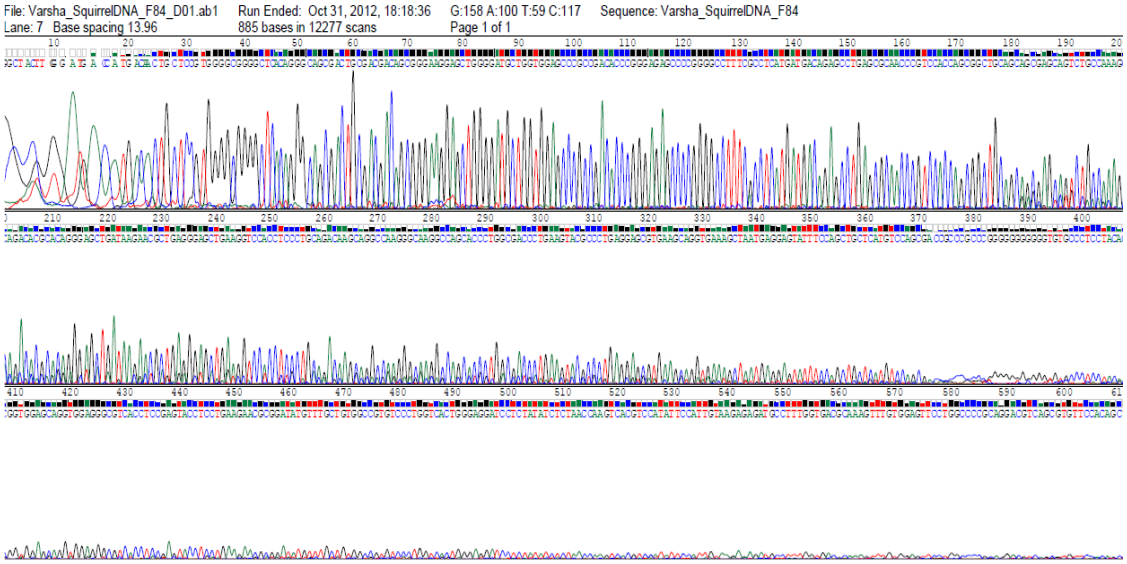
A2. Alignment of C57BL/6 (lab bl6) and wild mouse (lab wmus) sequence with GenBank sequences for C57BL/6 (ncbibl6) and *R.norvegicus* (ncbirattus)



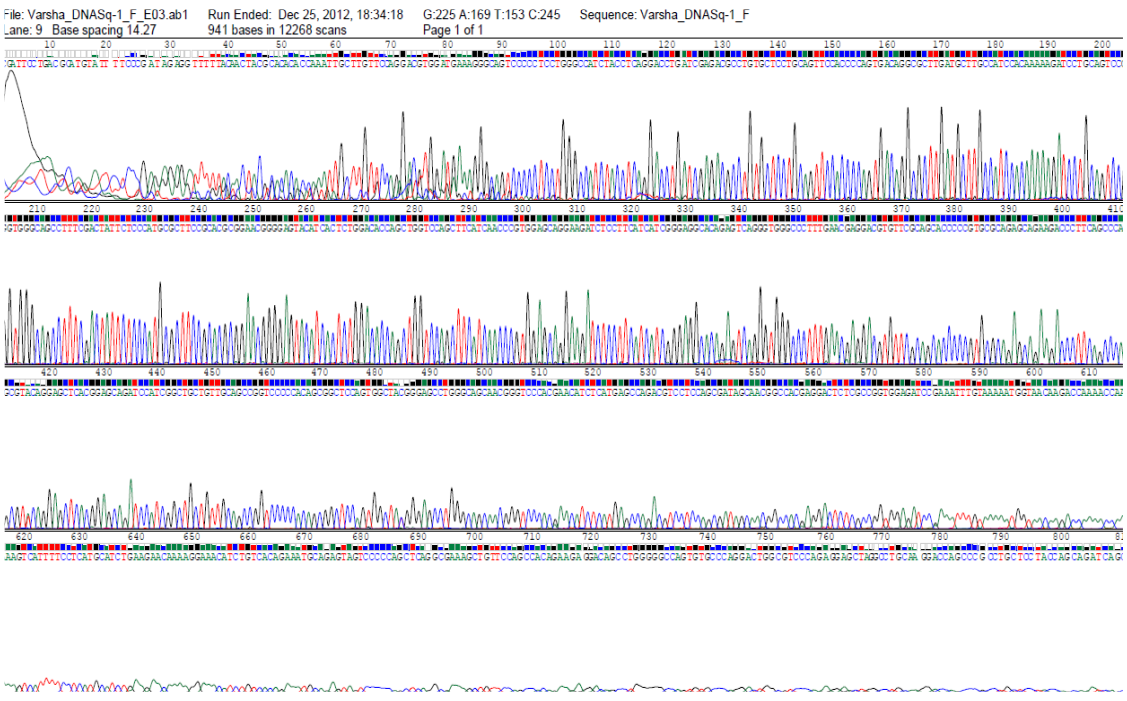
A3. Alignment of sequences flanking the *Per2* 'start' codon in mammals with that of *F.palmarum*



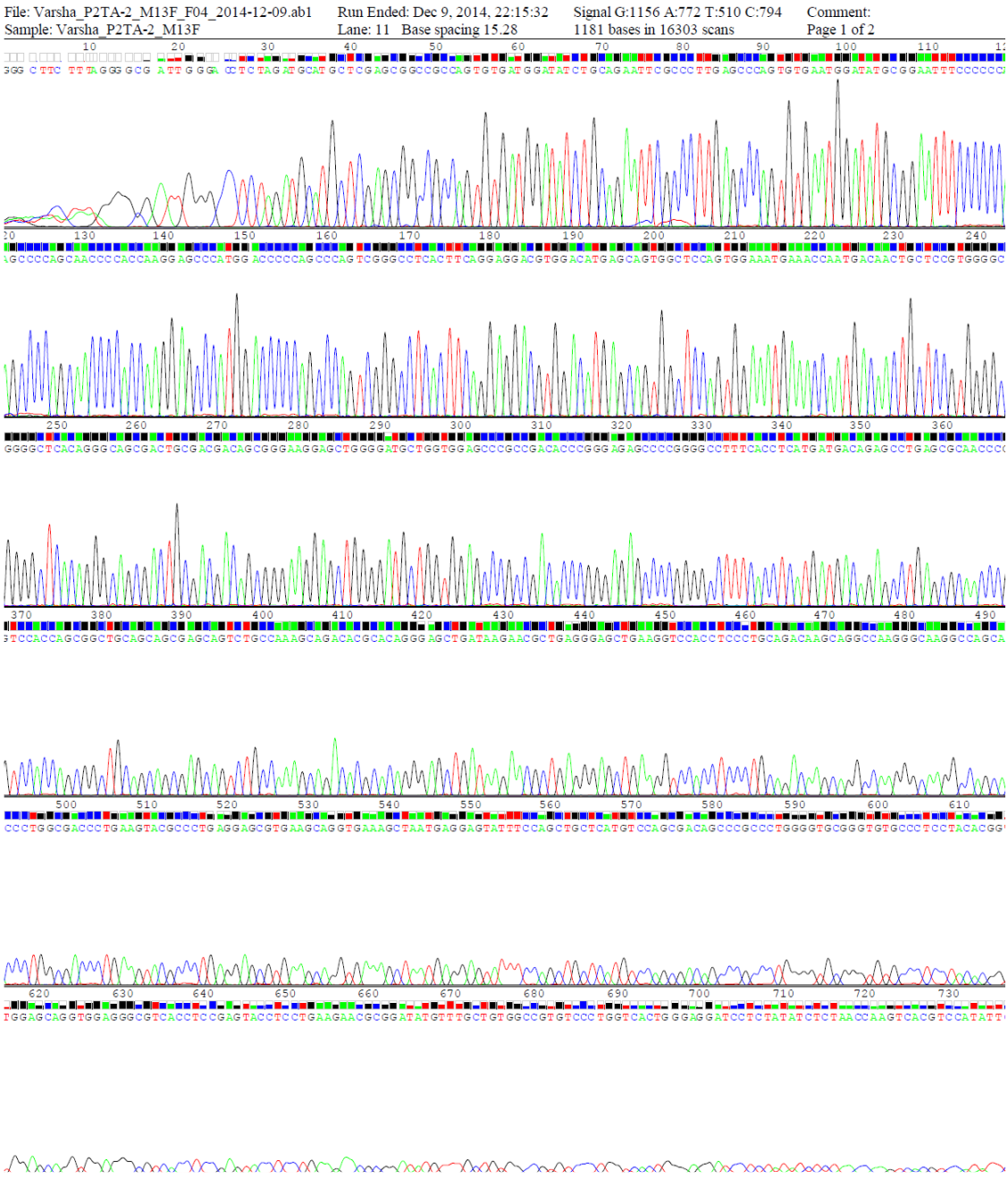
C5. Sequencing PCR amplicon eluate F84/R1146. Primer F84.



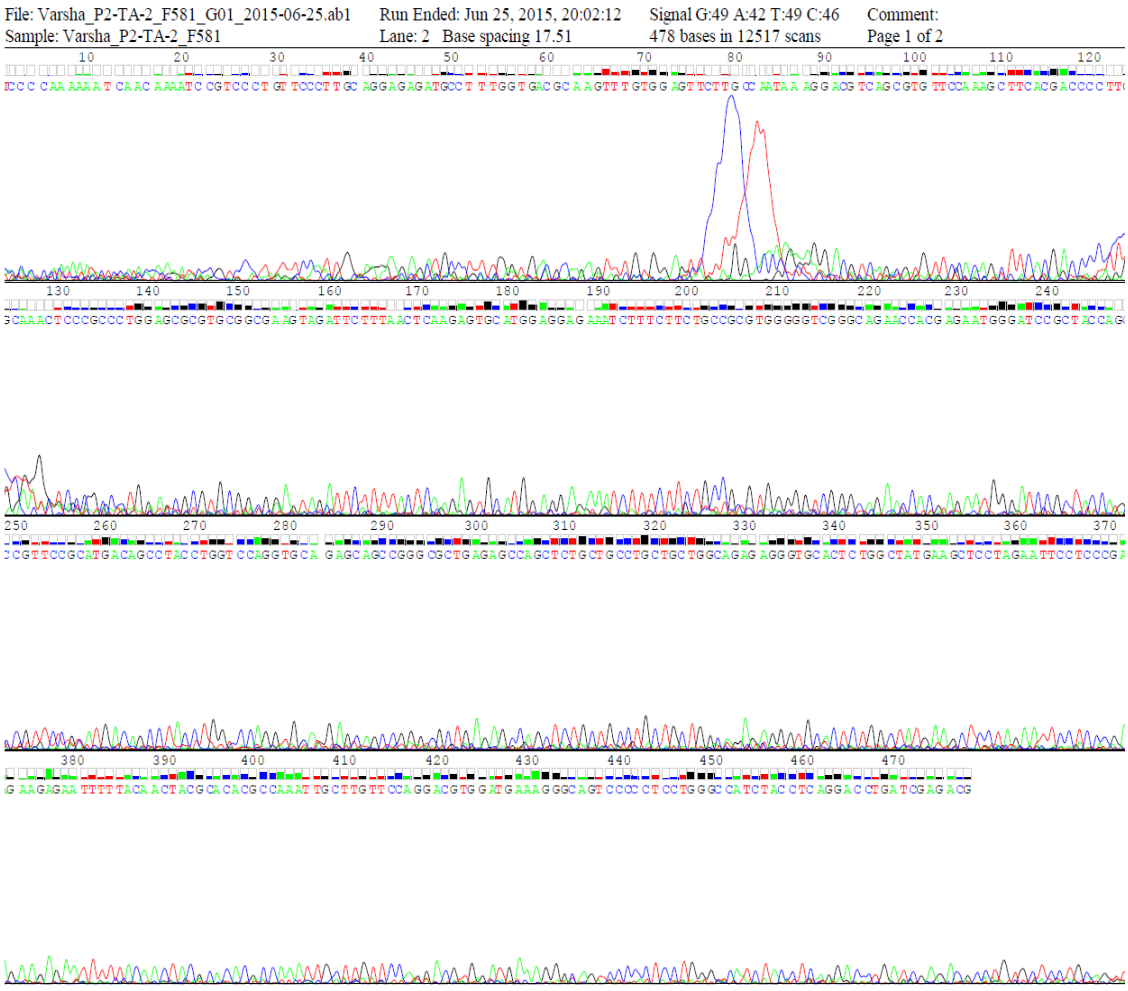
C6. Sequencing PCR amplicon eluate F927/R2216. Primer F927



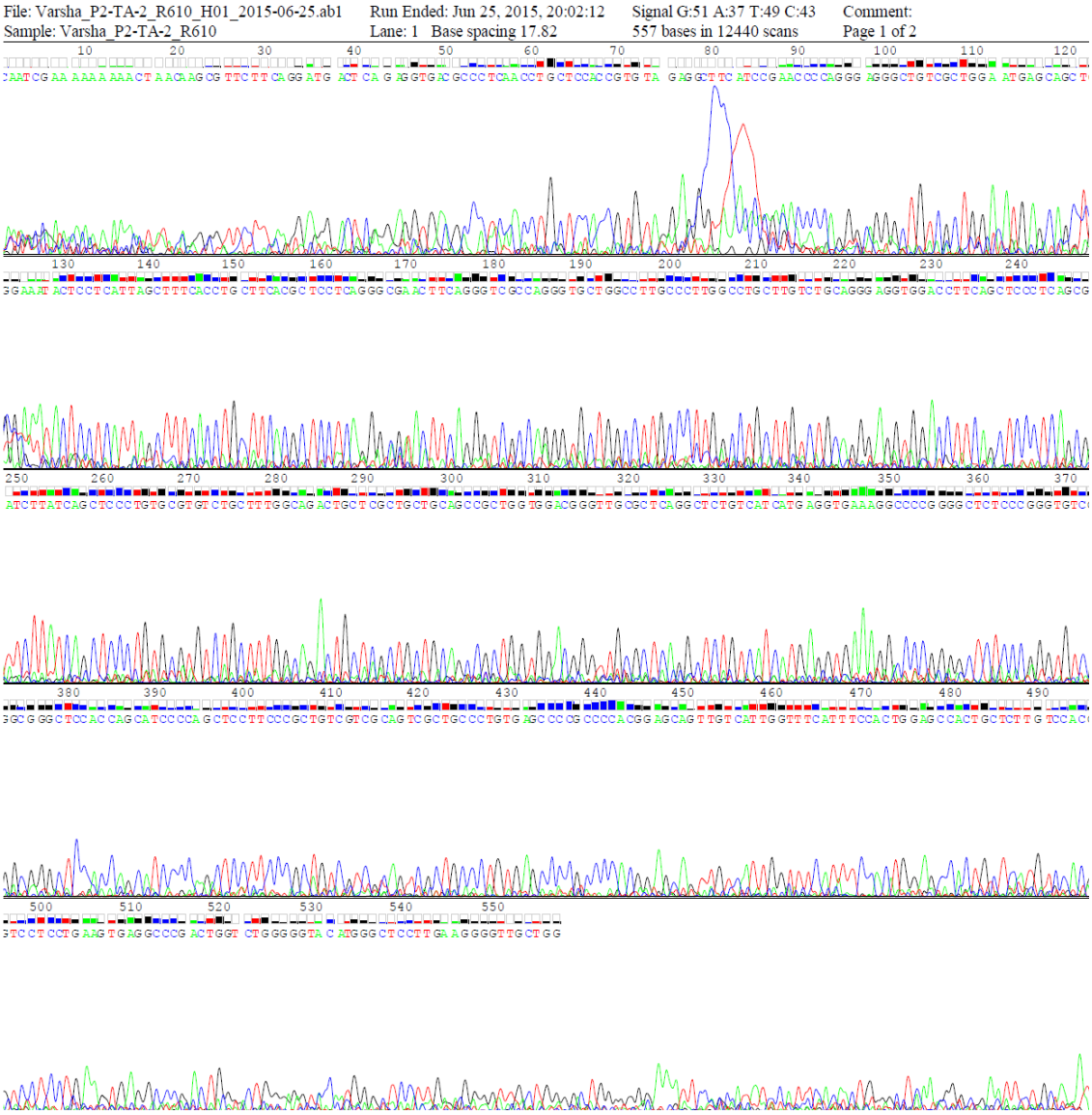
C7. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert. Primer M13F



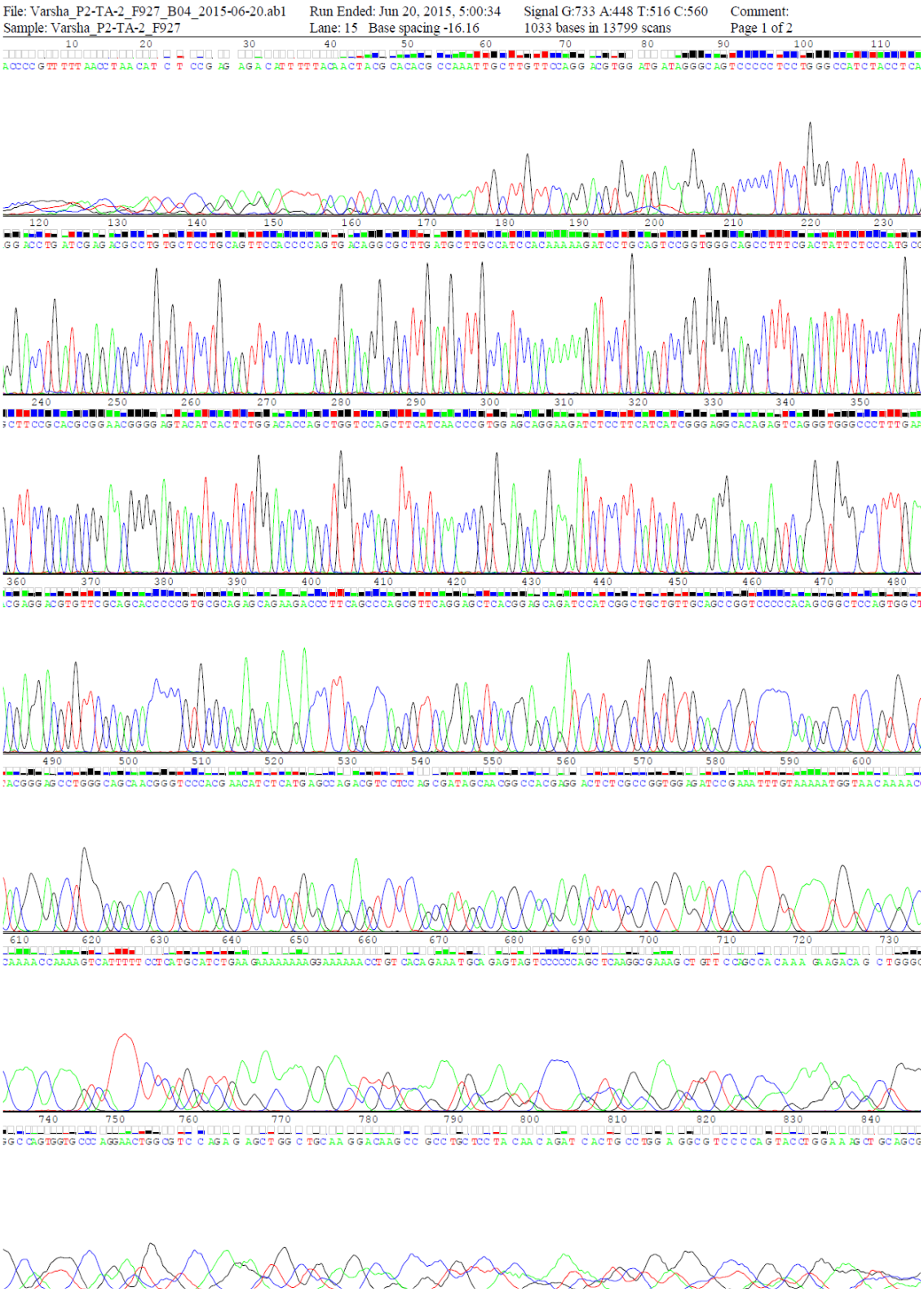
C8. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert. Primer F581



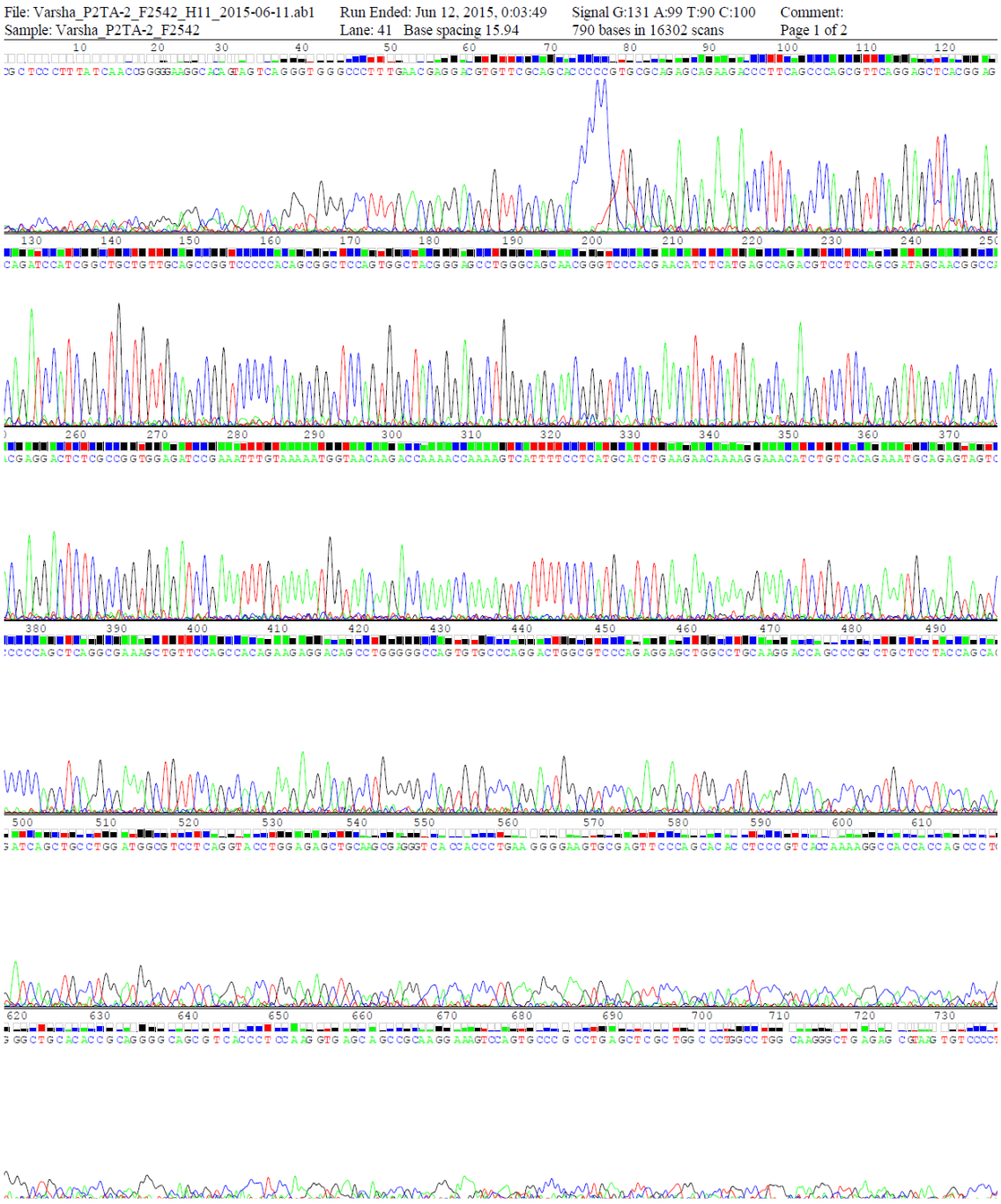
C9. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert. Primer R610



C 10. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert. Primer F927

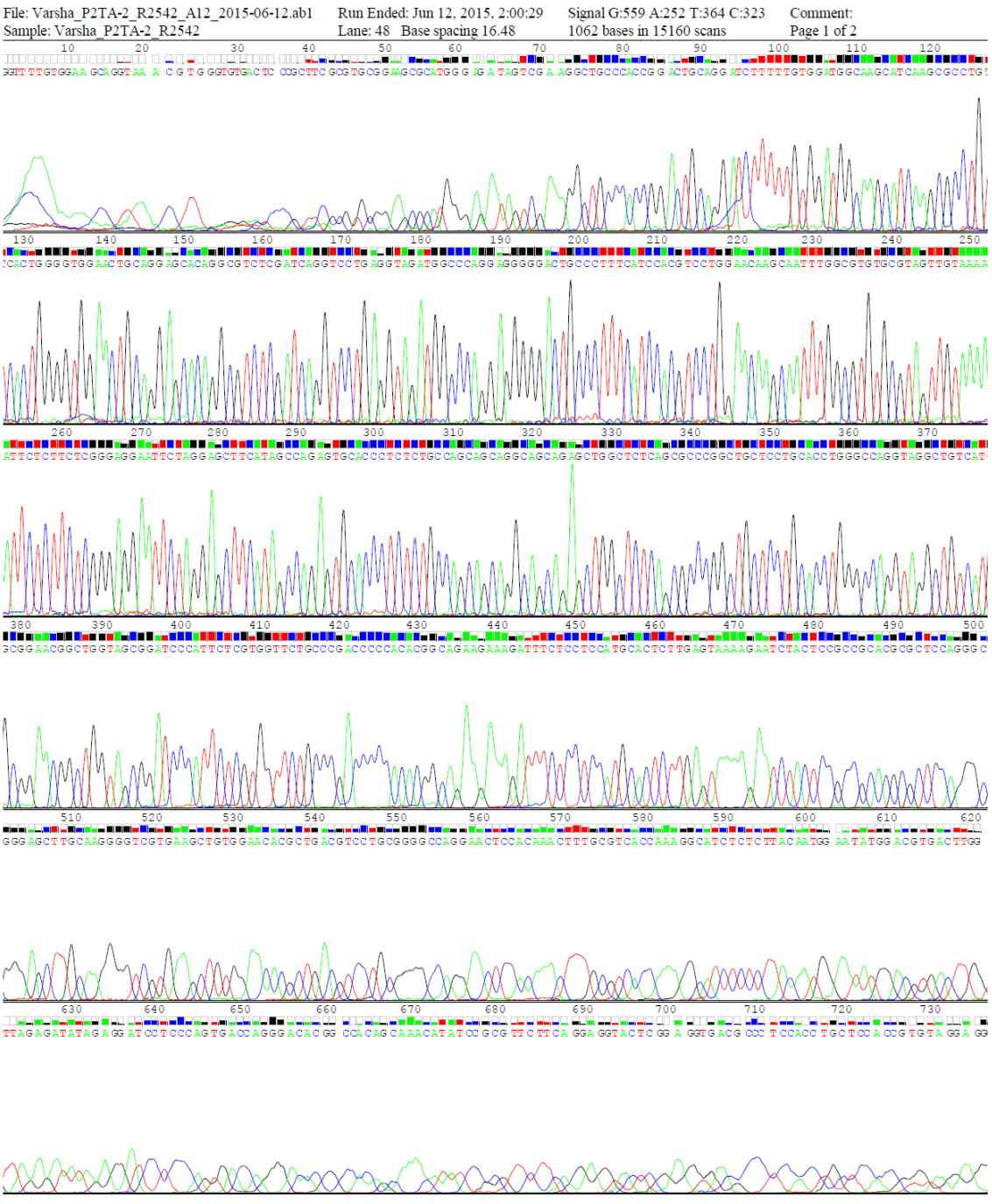


C11. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert. Primer F2542



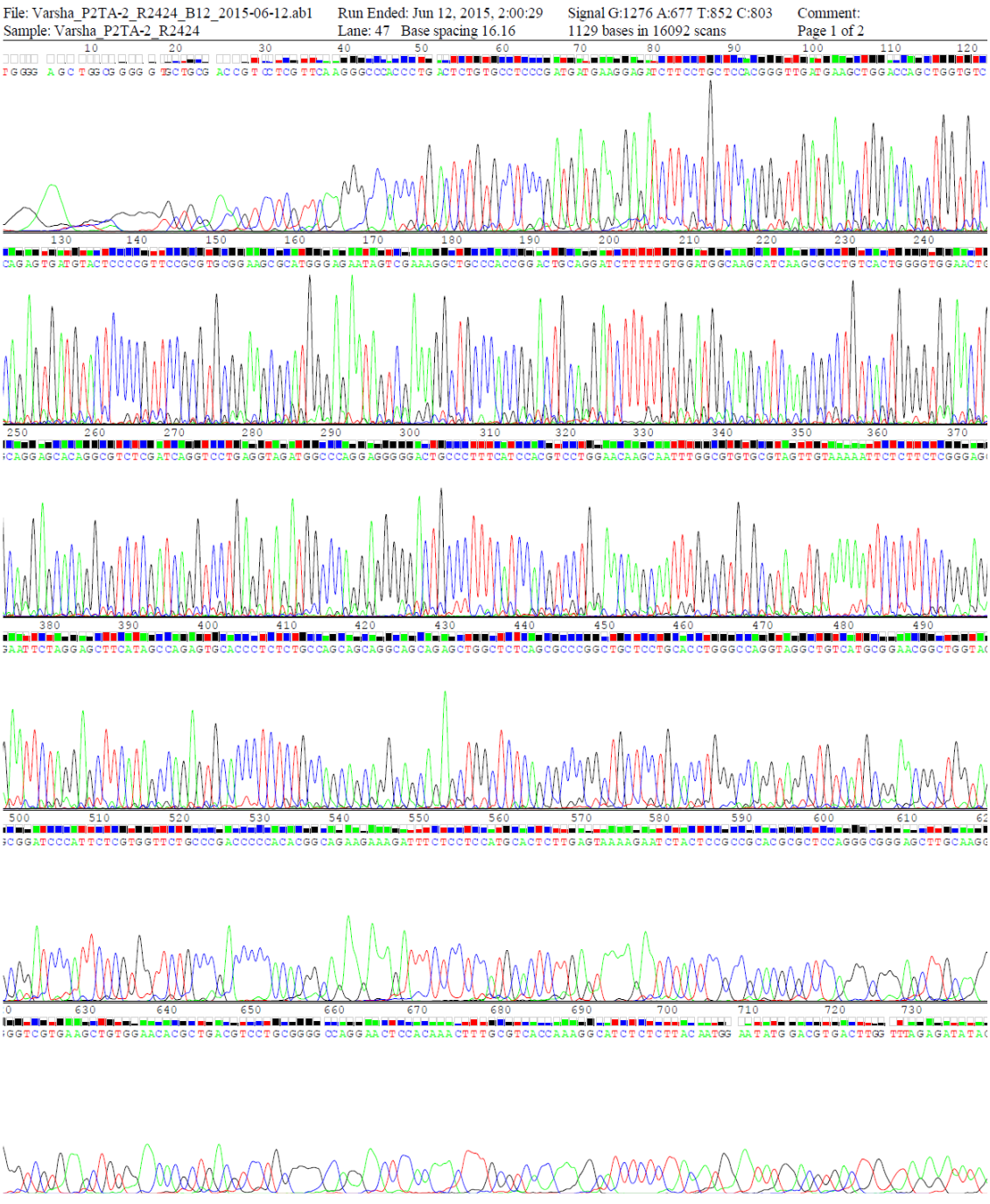
C 12. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer R2542



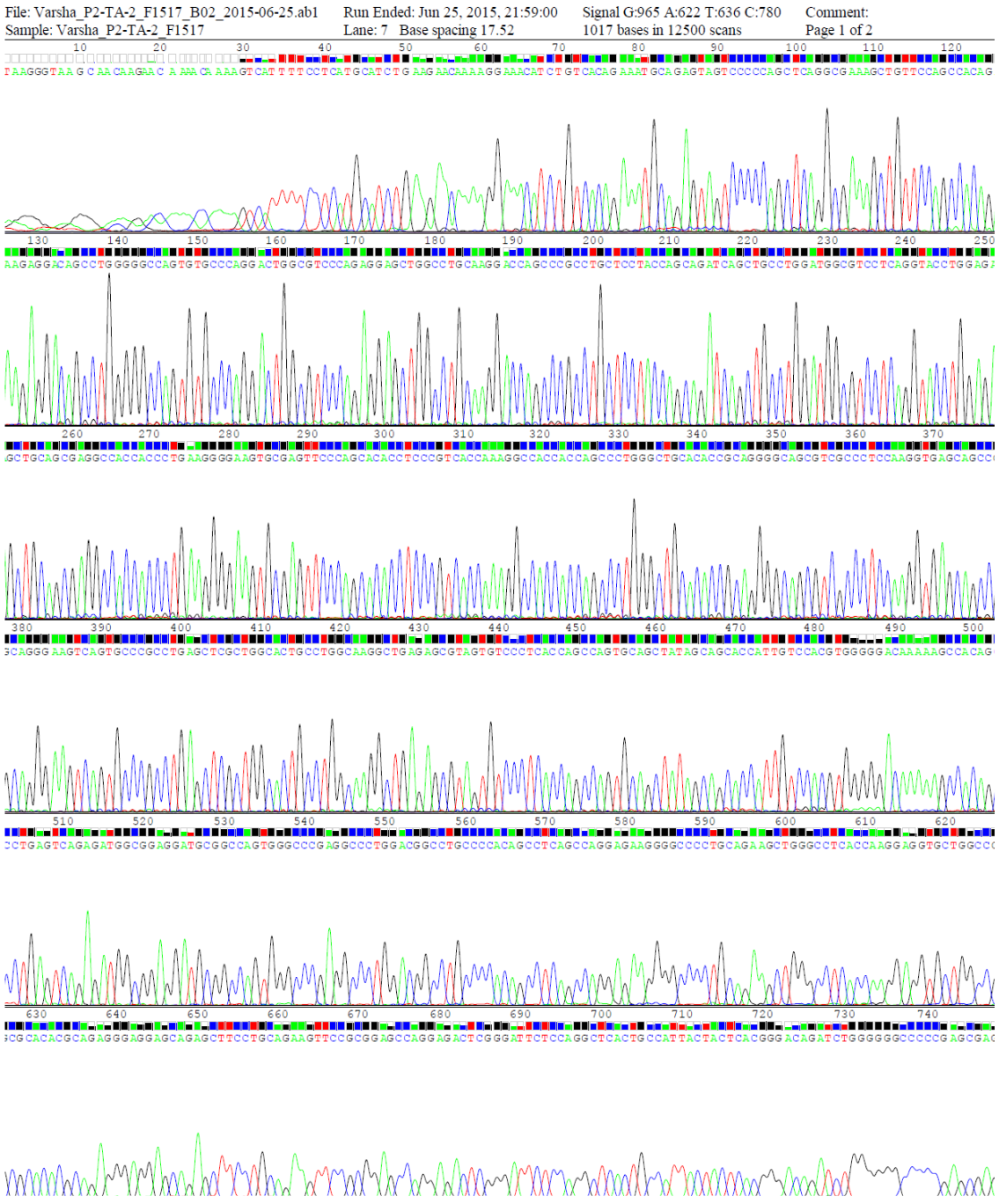
C 13. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer. R2424



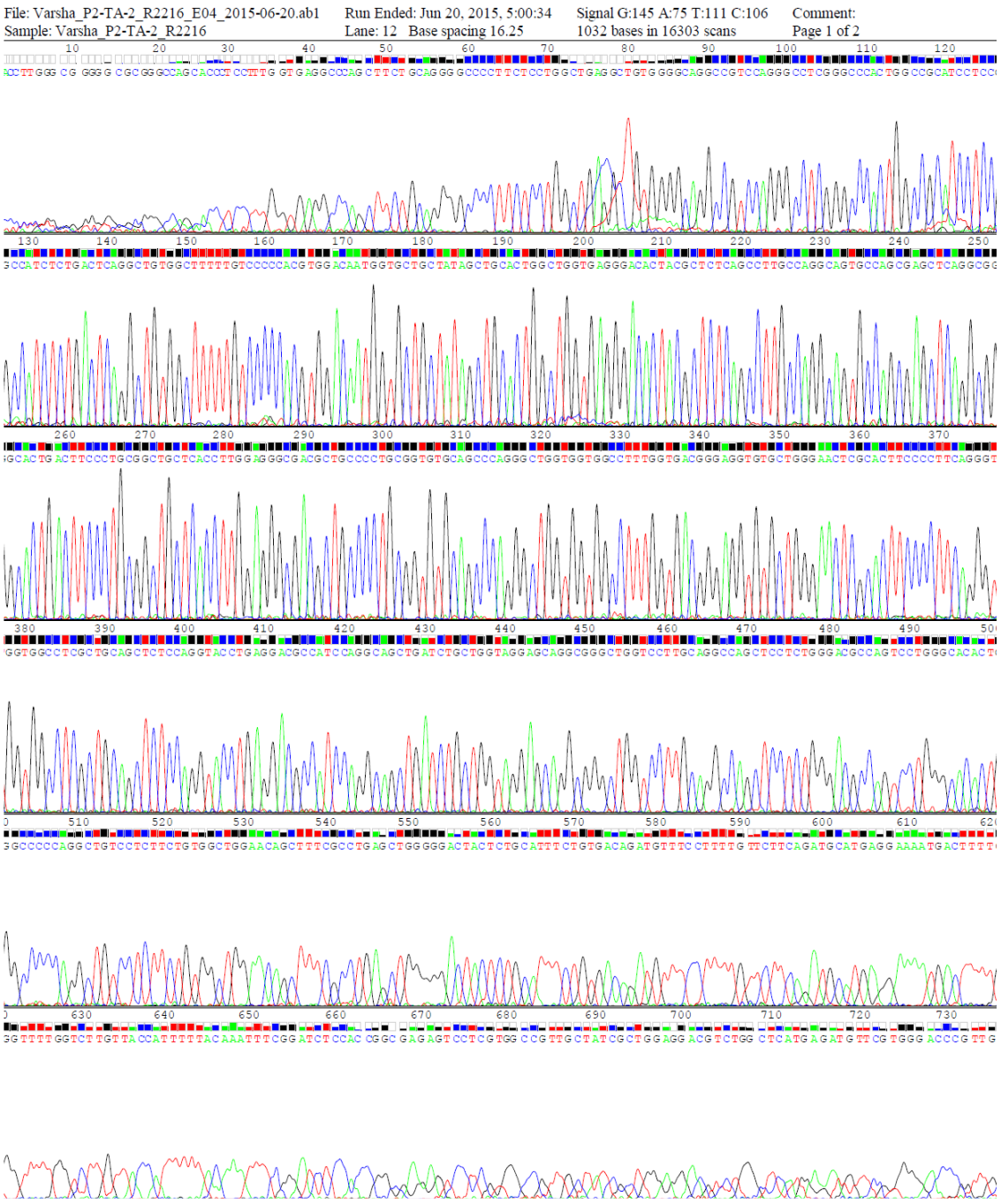
C 14. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer. F1517



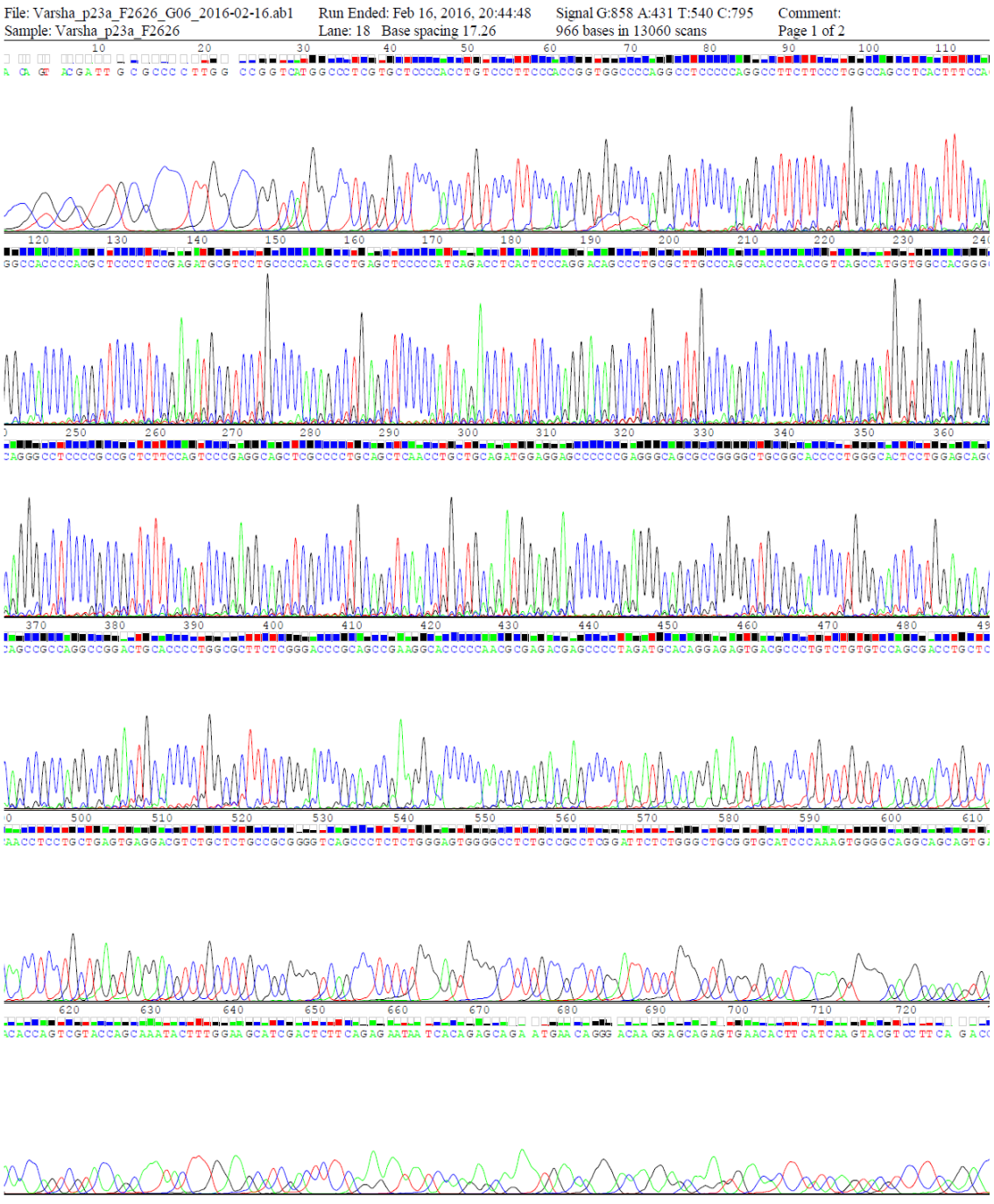
C 15. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer R2216



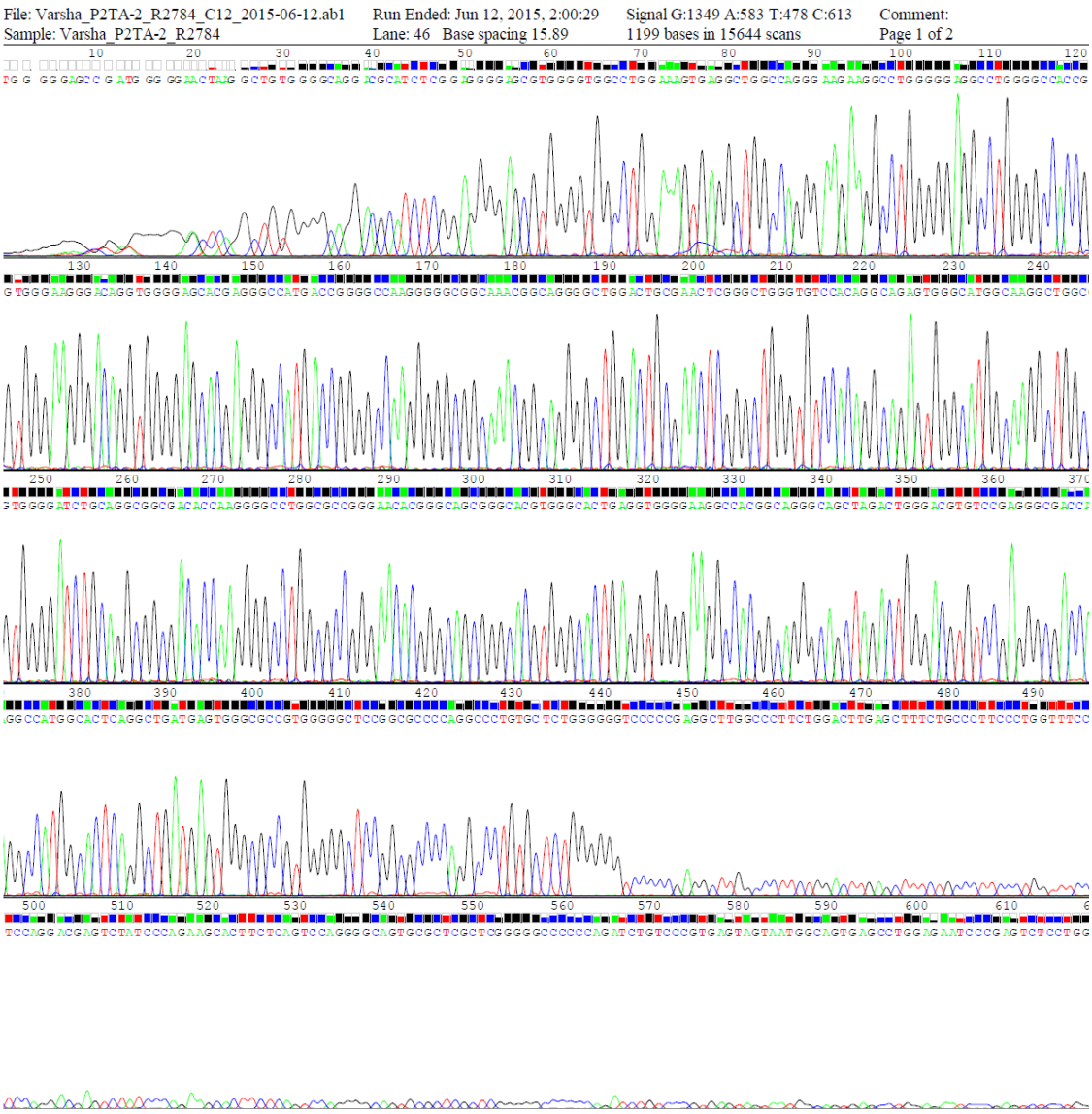
C 16. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer F2626



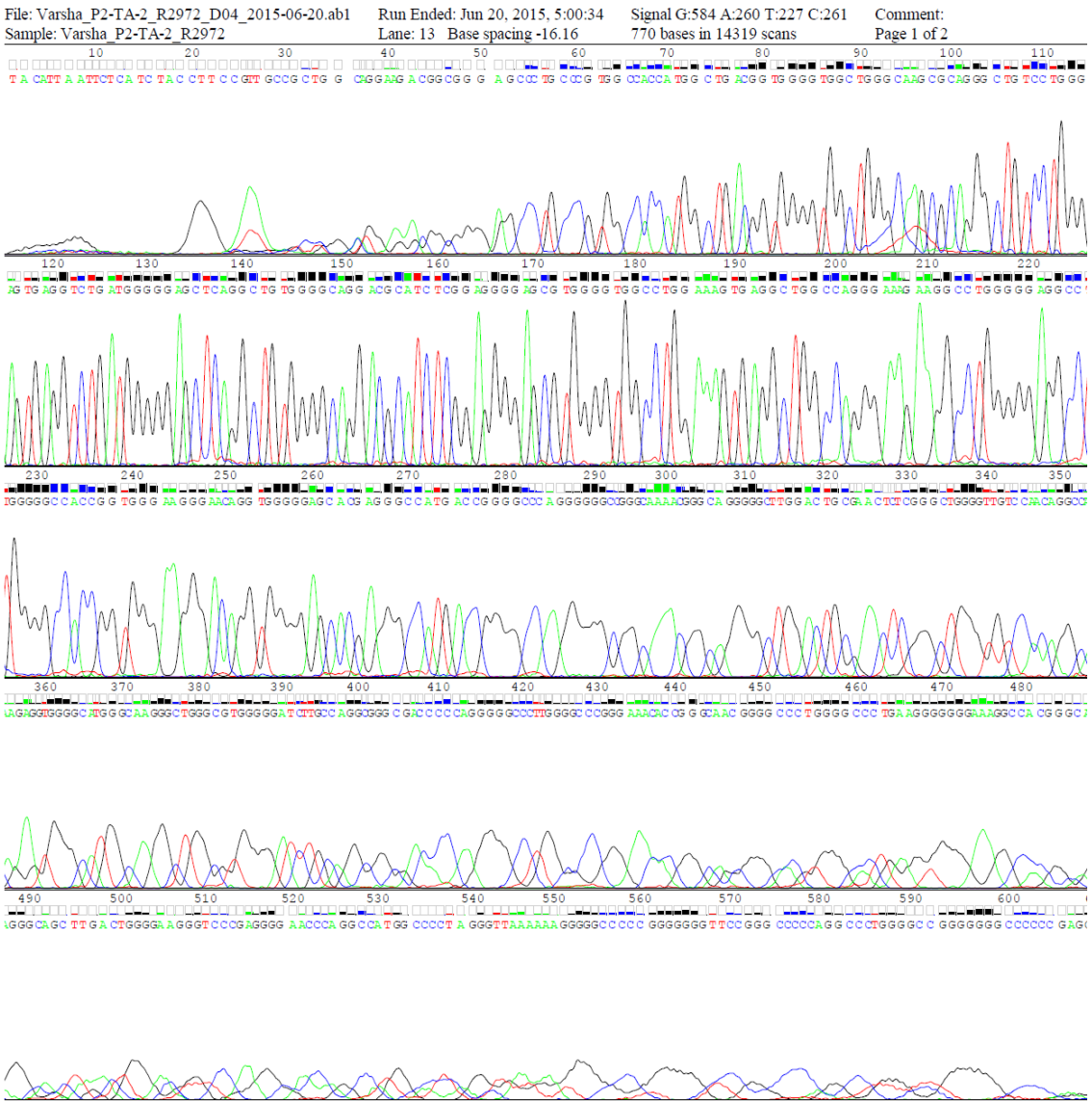
C 17. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer R2784.



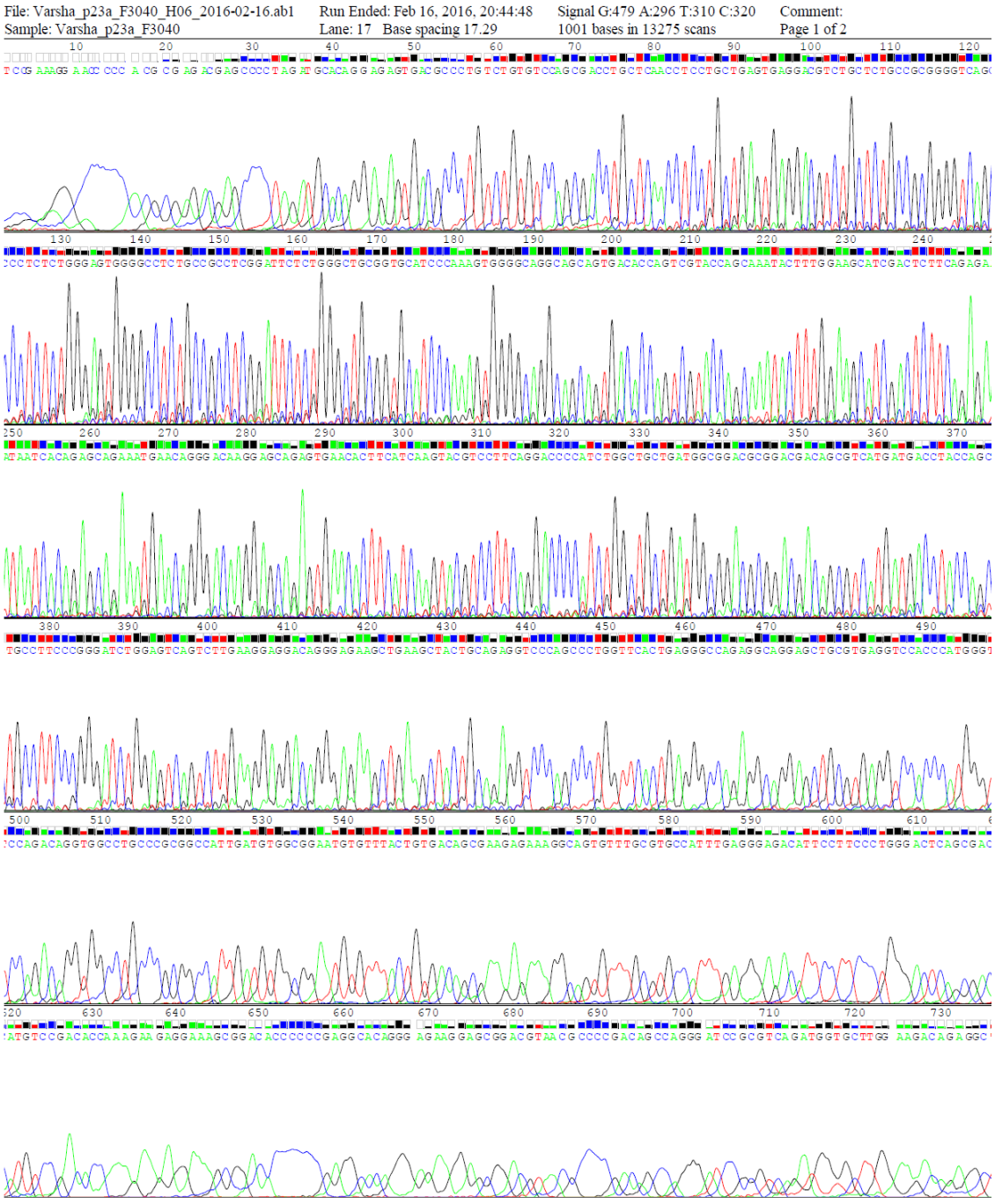
C 18. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer R2972



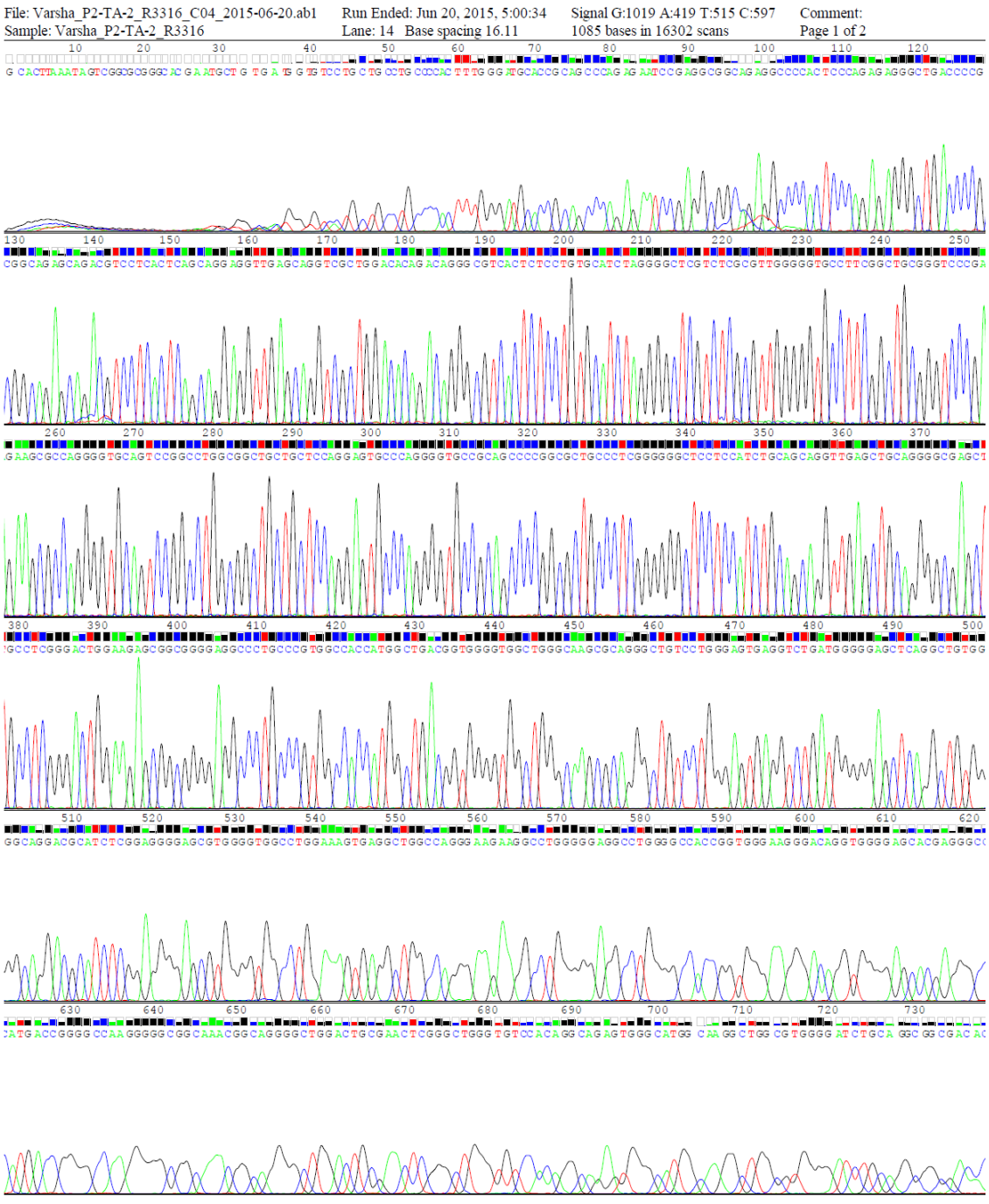
C 19. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer F3040



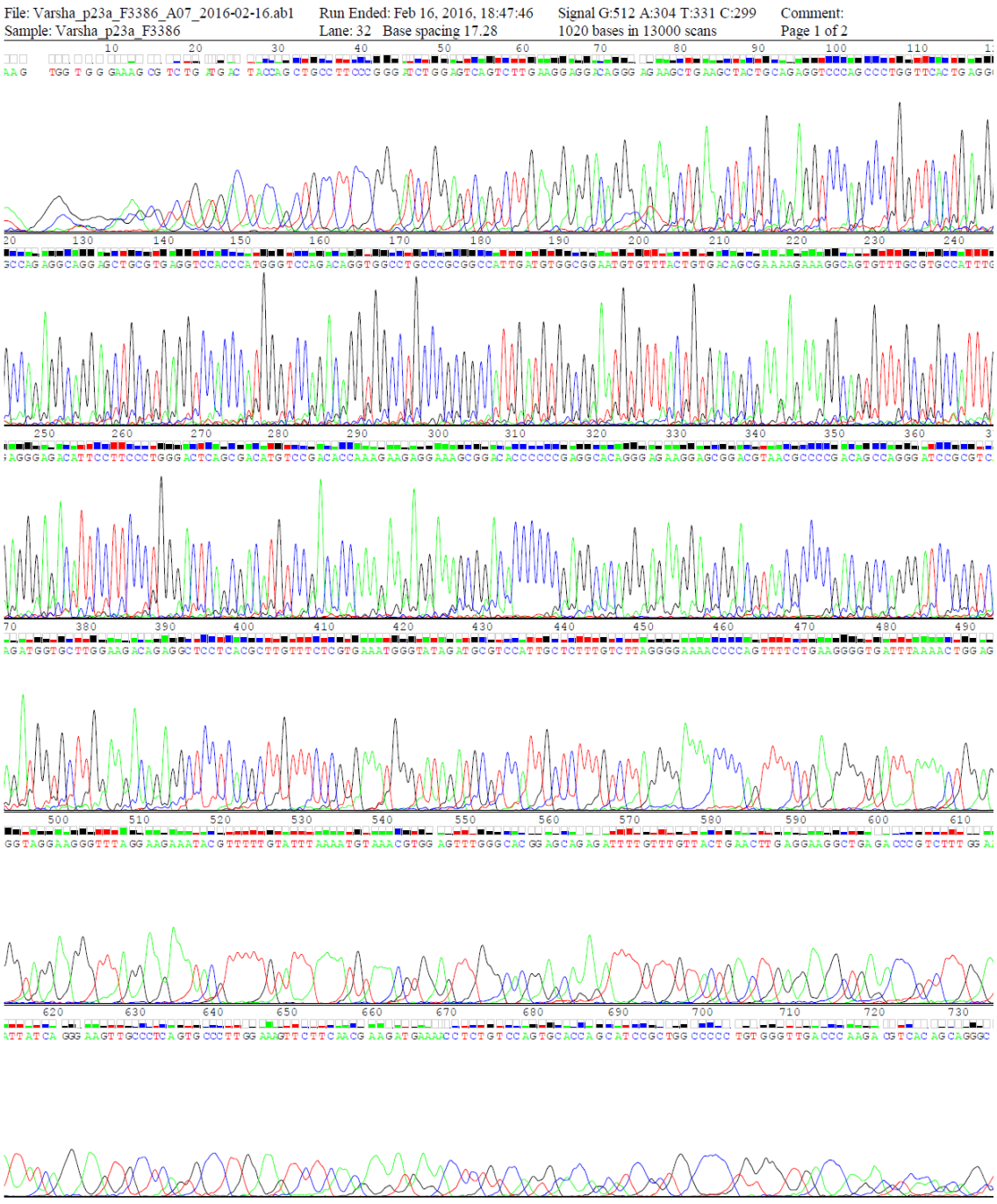
C 20. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer R3316.



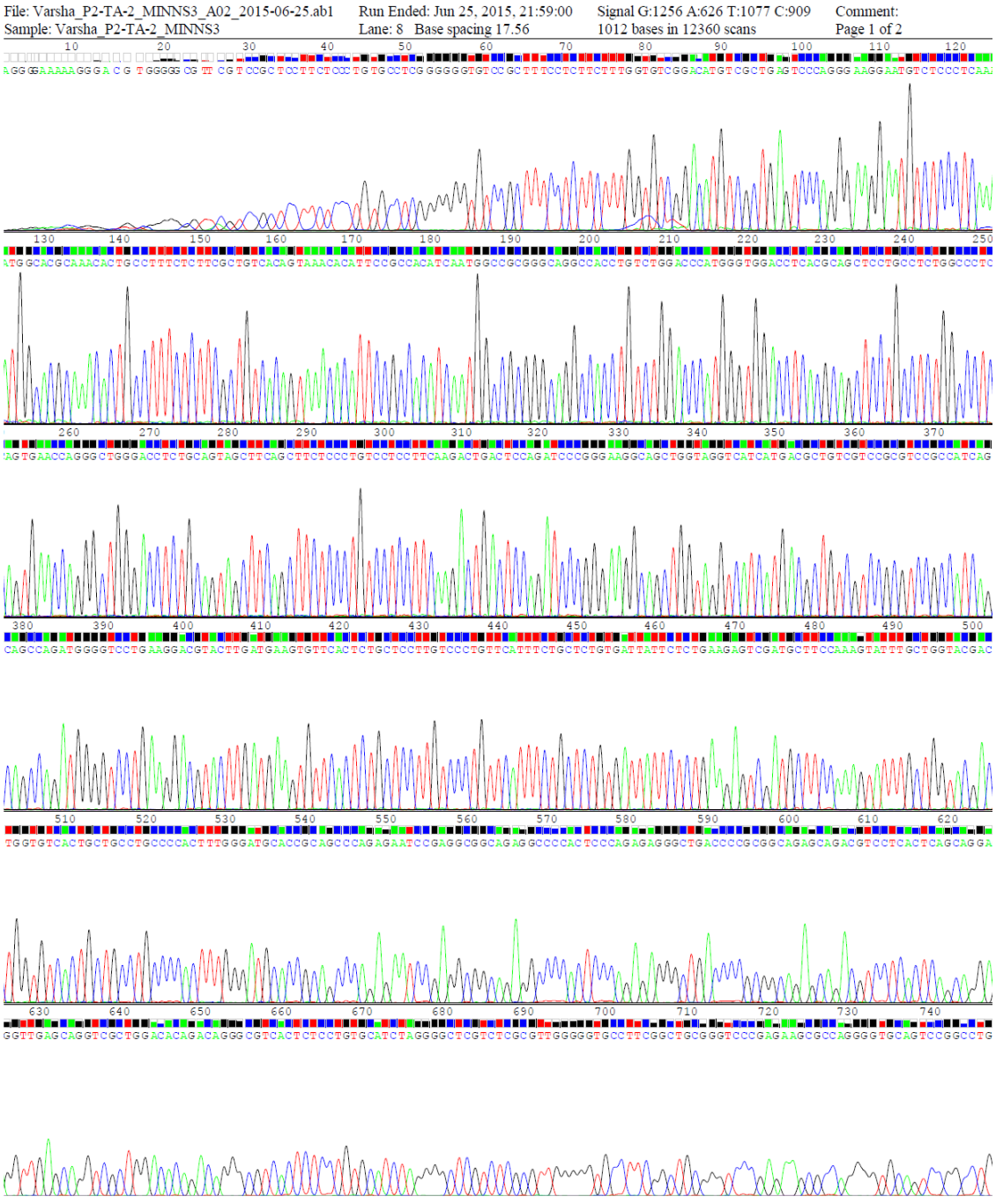
C 21. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer F3386



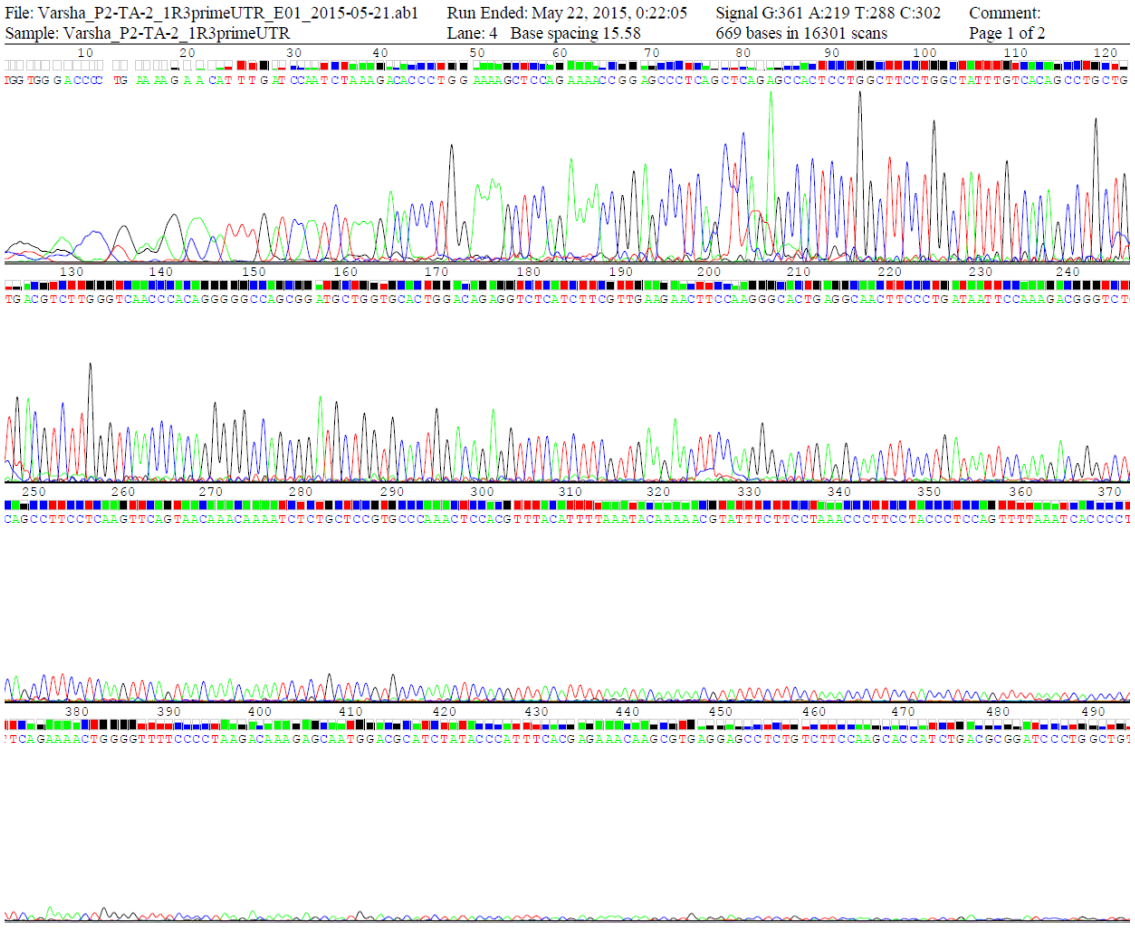
C 22. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer minus 3R3'UTR



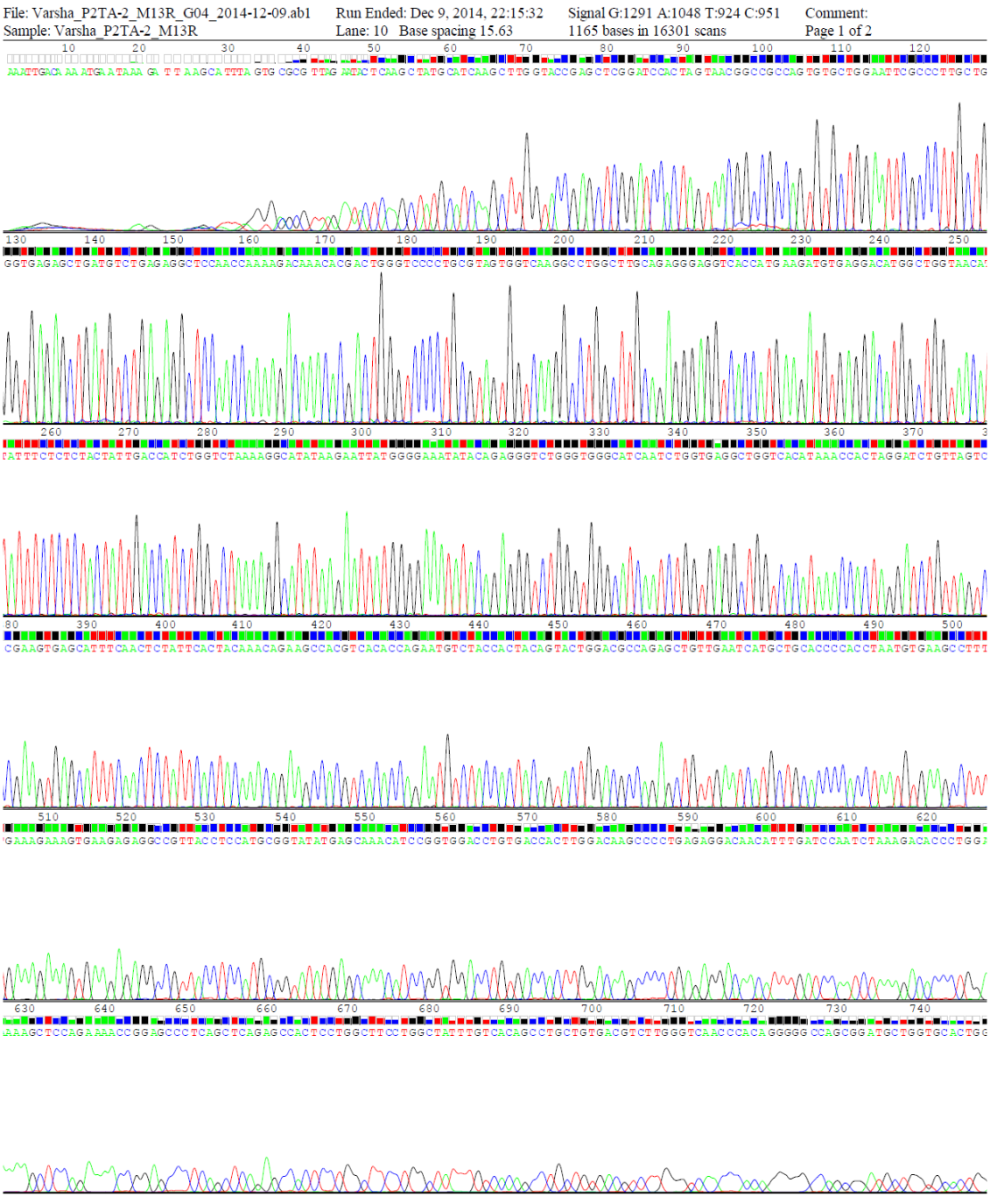
C 23. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer minus1 R3'UTR



C 24. Sequencing of TA-XL plasmid with *F. palmarum Per2* CDS insert.

Primer M13R



D 3. Complete nucleotide sequence of *F. palmarum Per2* derived from C7 to C 24

GTGAATGGATATGCGGAATTTCCCCCCAGCCCCAGCAACCCCACCAAGGA
GCCCATGGACCCCCAGCCCAGTCGGGCCTCACTTCAGGAGGACGTGGACAT
GAGCAGTGGCTCCAGTGGAAATGAAACCAATGACAACTGCTCCGTGGGGC
GGGGCTCACAGGGCAGCGACTGCGACGACAGCGGGAAGGAGCTGGGGATG
CTGGTGGAGCCCGCCGACACCCGGGAGAGCCCCGGGGCCTTTCACCTCATG
ATGACAGAGCCTGAGCGCAACCCGTCCACCAGCGGCTGCAGCAGCGAGCA
GTCTGCCAAAGCAGACACGCACAGGGAGCTGATAAGAACGCTGAGGGAGC
TGAAGGTCCACCTCCCTGCAGACAAGCAGGCCAAGGGCAAGGCCAGCACC
CTGGCGACCCTGAAGTACGCCCTGAGGAGCGTGAAGCAGGTGAAAGCTAA
TGAGGAGTATTTCCAGCTGCTCATGTCCAGCGACAGCCCGCCCTGGGGTGC
GGGTGTGCCCTCCTACACGGTGGAGCAGGTGGAGGGCGTCACCTCCGAGTA
CCTCCTGAAGAACGCGGATATGTTTGCTGTGGCCGTGTCCCTGGTCACTGGG
AGGATCCTCTATATCTCTAACCAAGTCACGTCCATATTCCATTGTAAGAGAG
ATGCCTTTGGTGACGCAAAGTTTGTGGAGTTCCTGGGCCCGCAGGACGTCA
GCGTGTTCCACAGCTTCACGACCCCTTGCAAGCTCCCGCCCTGGAGCGCGT
GCGGCGGAGTAGATTCTTTTACTCAAGAGTGCATGGAGGAGAAATCTTTTT
TCTGCCGTGTGGGGGTCGGGCAGAACCACGAGAATGGGATCCGCTACCAGC
CGTTCCGCATGACAGCCTACCTGGCCCAGGTGCAGGAGCAGCCGGGCGCTG
AGAGCCAGCTCTGCTGCCTGCTGCTGGCAGAGAGGGTGCACTCTGGCTATG
AAGCTCCTAGAATTCCTCCCGAGAAGAGAATTTTTACAACTACGCACACGC
CAAATTGCTTGTTCCAGGACGTGGATGAAAGGGCAGTCCCCCTCCTGGGCC
ATCTACCTCAGGACCTGATCGAGACGCCTGTGCTCCTGCAGTTCCACCCCAG
TGACAGGCGCTTGATGCTTGCCATCCACAAAAAGATCCTGCAGTCCGGTGG
GCAGCCTTTCGACTATTCTCCCATGCGCTTCCGCACGCGGAACGGGGAGTA
CATCACTCTGGACACCAGCTGGTCCAGCTTCATCAACCCGTGGAGCAGGAA
GATCTCCTTCATCATCGGGAGGCACAGAGTCAGGGTGGGCCCTTTGAACGA
GGACGTGTTCGCAGCACCCCCGTGCGCAGAGCAGAAGACCCTTCAGCCCAG
CGTTCAGGAGCTCACGGAGCAGATCCATCGGCTGCTGTTGCAGCCGGTCCC
CCACAGCGGCTCCAGTGGCTACGGGAGCCTGGGCAGCAACGGGTCCCACG
AACATCTCATGAGCCAGACGTCCTCCAGCGATAGCAACGGCCACGAGGACT
CTCGCCGGTGGAGATCCGAAATTTGTAAAAATGGTAACAAGACCAAAACCA
AAAGTCATTTTCCTCATGCATCTGAAGAACAAAAGGAAACATCTGTCACAG
AAATGCAGAGTAGTCCCCCAGCTCAGGCGAAAGCTGTTCCAGCCACAGAAG
AGGACAGCCTGGGGGCCAGTGTGCCCAGGACTGGCGTCCCAGAGGAGCTG
GCCTGCAAGGACCAGCCCGCCTGCTCCTACCAGCAGATCAGCTGCCTGGAT
GGCGTCCTCAGGTACCTGGAGAGCTGCAGCGAGGCCACCACCCTGAAGGG
GAAGTGCGAGTTCCCAGCACACCTCCCGTCACCAAAGGCCACCACCAGCCC
TGGGCTGCACACCGCAGGGGCAGCGTCGCCCTCCAAGGTGAGCAGCCGCA
GGGAAGTCAGTGCCCGCCTGAGCTCGCTGGCACTGCCTGGCAAGGCTGAGA
GCGTAGTGTCCCTCACCAGCCAGTGCAGCTATAGCAGCACCATTGTCCACG

TGGGGGACAAAAAGCCACAGCCTGAGTCAGAGATGGCGGAGGATGCGGCC
AGTGGGCCCGAGGCCCTGGACGGCCTGCCCCACAGCCTCAGCCAGGAGAA
GGGGCCCCTGCAGAAGCTGGGCCTCACCAAGGAGGTGCTGGCCGCGCACA
CGCAGAGGGAGGAGCAGAGCTTCCTGCAGAAGTTCCGCGGAGCCAGGAGA
CTCGGGATTCTCCAGGCTCACTGCCATTACTACTCACGGGACAGATCTGGG
GGGCCCCCGAGCGAGCGCACTGCCCCTGGACTGAGAAGTGCTTCTGGGATA
GACTCGTCCTGGAGGAAACCAGGGAAGGGCAGAAAGCTCAAGTCCAGAAG
GGCCAAGCCTCGGGGGACCCCCCAGAGCACAGGGCCTGGGGCGCCGGAGC
CCCCACGGCGCCCACTCATCAGCCTGAGTGCCATGGCCTGGTCGCCCTCGG
ACACGTCCCAGTCTAGCTGCCCTGCCGTGGCCTTCCCCACCTCAGTGCCCAC
GTGCCCGCTGCCCGTGTTCCCGGCGCCAGGCCCCTTGGTGTCGCCGCCTGCA
GATCCCCACGCCAGCCTTGCCATGCCCACTCTGCCTGTGGACACCCAGCCC
GAGTTCGCAGTCCAGCCCCTGCCGTTTGCCGCCCCCTTGGCCCCGGTCATGG
CCCTCGTGCTCCCCACCTGTCCCTTCCCACCGGTGGCCCCAGGCCTCCCCCA
GGCCTTCTTCCCTGGCCAGCCTCACTTTCCAGGCCACCCCACGCTCCCCTCC
GAGATGCGTCCTGCCCCACAGCCTGAGCTCCCCCATCAGACCTCACTCCCA
GGACAGCCCTGCGCTTGCCCAGCCACCCCACCGTCAGCCATGGTGGCCACG
GGCAGGGCCTCCCCGCCGCTCTTCCAGTCCCGAGGCAGCTCGCCCCTGCAG
CTCAACCTGCTGCAGATGGAGGAGCCCCCCGAGGGCAGCGCCGGGGCTGC
GGCACCCCTGGGCACTCCTGGAGCAGCAGCCGCCAGGCCGGACTGCACCCC
TGGCGCTTCTCGGGACCCGCAGCCGAAGGCACCCCCAACGCGAGACGAGCC
CCTAGATGCACAGGAGAGTGACGCCCTGTCTGTGTCCAGCGACCTGCTCAA
CCTCCTGCTGAGTGAGGACGTCTGCTCTGCCGCGGGGTCAGCCCTCTCTGGG
AGTGGGGCCTCTGCCGCCTCGGATTCTCTGGGCTGCGGTGCATCCCAAAGT
GGGGCAGGCAGCAGTGACACCAGTCGTACCAGCAAATACTTTGGAAGCATC
GACTCTTCAGAGAATAATCACAGAGCAGAAATGAACAGGGACAAGGAGCA
GAGTGAACACTTCATCAAGTACGTCCTTCAGGACCCCATCTGGCTGCTGAT
GGCGGACGCGGACGACAGCGTCATGATGACCTACCAGCTGCCTTCCCGGGA
TCTGGAGTCAGTCTTGAAGGAGGACAGGGAGAAGCTGAAGCTACTGCAGA
GGTCCCAGCCCTGGTTCACTGAGGGCCAGAGGCAGGAGCTGCGTGAGGTCC
ACCCATGGGTCCAGACAGGTGGCCTGCCCGCGGCCATTGATGTGGCGGAAT
GTGTTTACTGTGACAGCGAAAAGAAAGGCAGTGTTTGCGTGCCATTTGAGG
GAGACATTCCTTCCCTGGGACTCAGCGACATGTCCGACACCAAAGAAGAGG
AAAGCGGACACCCCCCGAGGCACAGGGAGAAGGAGCGGACG**TAA**CGCCC
CGACAGCCAGGGATCCGCGTCAGATGGTGCTTGGAAGACAGAGGCTCCTCA
CGCTTGTTTCTCGTGAAATGGGTATAGATGCGTCCATTGCTCTTTGTCTTAG
GGGAAAACCCCAGTTTTCTGAAGGGGTGATTTAAAACTGGAGGGTAGGAA
GGGTTTAGGAAGAAATACGTTTTTGTATTTAAAATGTAAACGTGGAGTTTG
GGCACGGAGCAGAGATTTTGTTTGTTACTGAACTTGAGGAAGGCTGAGACC
CGTCTTTGGAATTATCAGGGAAGTTGCCTCAGTGCCCTTGGAAGTTCTTCAA
CGAAGATGAGACCTCTGTCCAGTGCACCAGCATCCGCTGGCCCCCTGTGGG
TTGACCCAAGACGTCACAGCAGGCTGTGACAAATAGCCAGGAAGCCAGGA
GTGGCTCTGAGCTGAGGGCTCCGGTTTTCTGGAGCTTTTCCAGGGTGTCTTT

AGATTGGATCAAATGTTGTCCTCTCAGGGGCTTGTCCAAGTGGTCACAGGT
CCACCGGATGTTTGCTCATATACCGCATGGAGGTAACGGCCTCTCTTCACTT
TCTTTCAAAGGCTTCACATTAGGTGGGGTGCAGCATGATTCAACAGCTCTG
GCGTCCAGTACTGTAGTGGTAGACATTCTGGTGTGACGTGGCTTCTGTTTGT
AGTGAATAGAGTTGAAATGCTCACTTCGGACTAACAGATCCTAGTGGTTTA
TGTGACCAGCCTCACCAGATTGATGCCCACCCAGACCCTCTGTATATTTCCC
CATAATTCTTATATGCCTTTTAGACCAGATGGTCAATAGTAGAGAGAAATA
TGTTACCAGCCATGTCCTCACATCTTCATGGTGACCTCCCTCTGCAAGCCAG
GCCTTGACCACTACGCAGGGGACCCAGTCGTGTTTGTCTTTTGGTTGGAGCC
TCTCAGACATCAGCTCTCACCCAGC

Translate Tool - Results of translation

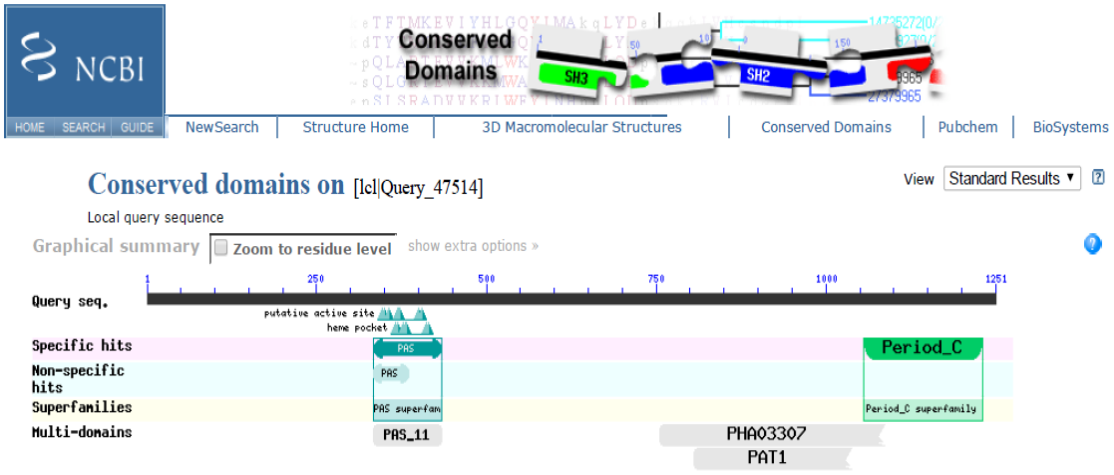
Open reading frames are highlighted in red.

5'3'Frame1

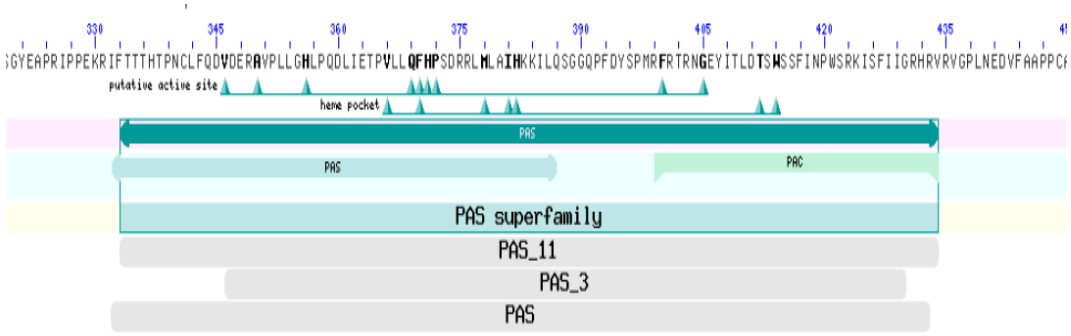
V N G Y A E F P P S P S N P T K E P **Met** D P Q P S R A S L Q E D V D **Met** S S G
S S G N E T N D N C S V G R G S Q G S D C D D S G K E L G **Met** L V E P A D T
R E S P G A F H L **Met** **Met** T E P E R N P S T S G C S S E Q S A K A D T H R E L
I R T L R E L K V H L P A D K Q A K G K A S T L A T L K Y A L R S V K Q V K
A N E E Y F Q L L **Met** S S D S P P W G A G V P S Y T V E Q V E G V T S E Y L L
K N A D **Met** F A V A V S L V T G R I L Y I S N Q V T S I F H C K R D A F G D A
K F V E F L G P Q D V S V F H S F T T P C K L P P W S A C G G V D S F T Q E
C **Met** E E K S F F C R V G V G Q N H E N G I R Y Q P F R **Met** T A Y L A Q V Q
E Q P G A E S Q L C C L L L A E R V H S G Y E A P R I P P E K R I F T T T H T P
N C L F Q D V D E R A V P L L G H L P Q D L I E T P V L L Q F H P S D R R
L **Met** L A I H K K I L Q S G G Q P F D Y S P **Met** R F R T R N G E Y I T L D T S
W S S F I N P W S R K I S F I I G R H R V R V G P L N E D V F A A P P C A E Q K
T L Q P S V Q E L T E Q I H R L L L Q P V P H S G S S G Y G S L G S N G S H E
H L **Met** S Q T S S S D S N G H E D S R R W R S E I C K N G N K T K T K S H F P
H A S E E Q K E T S V T E **Met** Q S S P P A Q A K A V P A T E E D S L G A S V P
R T G V P E E L A C K D Q P A C S Y Q Q I S C L D G V L R Y L E S C S E A T T
L K G K C E F P A H L P S P K A T T S P G L H T A G A A S P S K V S S R R E V
S A R L S S L A L P G K A E S V V S L T S Q C S Y S S T I V H V G D K K P Q P
E S E **Met** A E D A A S G P E A L D G L P H S L S Q E K G P L Q K L G L T K E V
L A A H T Q R E E Q S F L Q K F R G A R R L G I L Q A H C H Y Y S R D R S G
G P P S E R T A P G L R S A S G I D S S W R K P G K G R K L K S R R A K P R G
T P Q S T G P G A P E P P R R P L I S L S A **Met** A W S P S D T S Q S S C P A V A
F P T S V P T C P L P V F P A P G P L V S P P A D P H A S L A **Met** P T L P V D T
Q P E F A V Q P L P F A A P L A P V **Met** A L V L P T C P F P P V A P G L P Q A
F F P G Q P H F P G H P T L P S E **Met** R P A P Q P E L P H Q T S L P G Q P C A C
P A T P P S A **Met** V A T G R A S P P L F Q S R G S S P L Q L N L L Q **Met** E E P
P E G S A G A A A P L G T P G A A A A R P D C T P G A S R D P Q P K A P P T R
D E P L D A Q E S D A L S V S S D L L N L L L S E D V C S A A G S A L S G S G

A S A A S D S L G C G A S Q S G A G S S D T S R T S K Y F G S I D S S E N N H
R A E **Met** N R D K E Q S E H F I K Y V L Q D P I W L L **Met** A D A D D S
V **Met** **Met** T Y Q L P S R D L E S V L K E D R E K L K L L Q R S Q P W F T E G
Q R Q E L R E V H P W V Q T G G L P A A I D V A E C V Y C D S E K K G S V C
V P F E G D I P S L G L S D **Met** S D T K E E E S G H P P R H R E K E R T **Stop** R
P D S Q G S A S D G A W K T E A P H A C F S **Stop** N G Y R C V H C S L
S **Stop** G K T P V F **Stop** R G D L K L E G R K G L G R N T F L Y L K C K R G V
W A R S R D F V C Y **Stop** T **Stop**G R L R P V F G I I R E V A S V P L E V L Q R
R **Stop** D L C P V H Q H P L A P C G L T Q D V T A G C D K **Stop** P G S Q E W
L **Stop** A E G S G F L E L F Q G V F R L D Q **Met** L S S Q G L V Q V V T G P P D
V C S Y T A W R **Stop** R P L F T F F Q R L H I R W G A A **Stop** F N S S G V Q Y C
S G R H S G V T W L L F V V N R V E **Met** L T S D **Stop** Q I L V V Y V T S L T R
L **Met** P T Q T L C I F P H N S Y **Met** P F R P D G Q **Stop** **Stop** R E I C Y Q P C P H
I F **Met** V T S L C K P G L D H Y A G D P V V F V F W L E P L R H Q L S P S

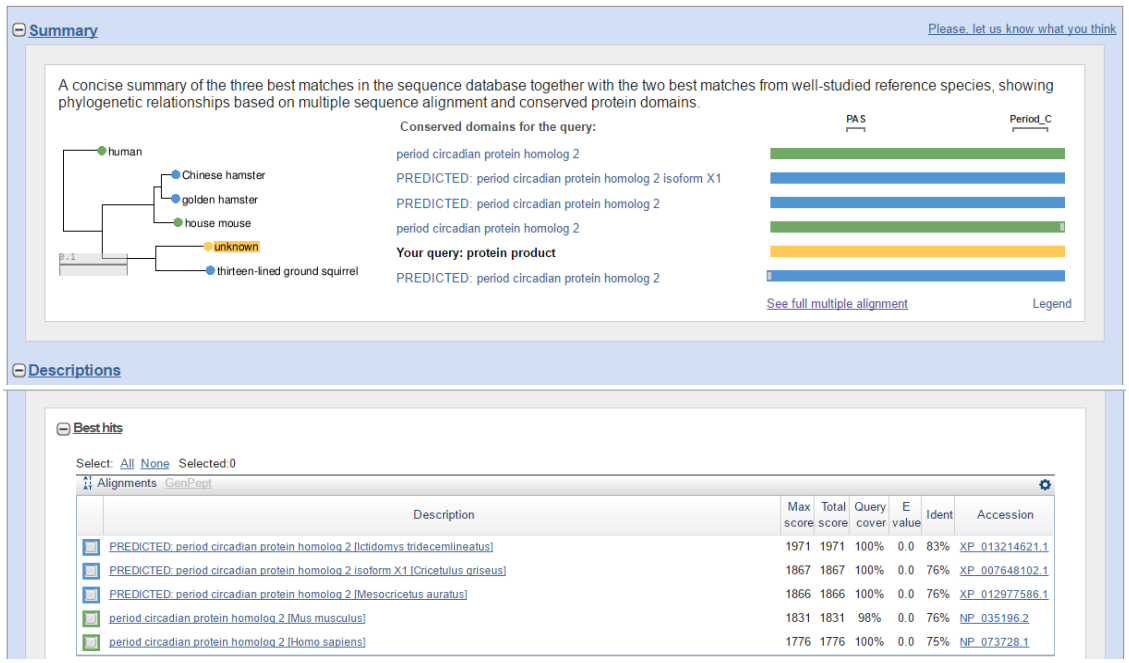
Conserved Domain Search and SMART BLAST in NCBI



The expanded view of the PAS domain region in the query



SMART BLAST - a snapshot of summary of the query sequence.



A4. CLUSTAL alignment of *F.palmarum* aminoacid sequence with that of *Mus musculus, Rattus norwegicus, Mesocricetus auratus, Cricetulus griseus, Ictidomys tridecemlineatus, Homo sapiens.The FASP domain is highlighted*

```
fpalper2          VNGYAEFPPSPSNPTKEPMDPQPSRASLQEDVDMSSGSSGNETNDNCSVGRGSQGSDCDD 60
ictidomysper2     MNGFAEFPPSTGDPTKELMDAQPSQALLQEDVDMSSGSSGN---ENCSMGRDSQGSDCED 57
hper2             MNGYAEFPPSPSNPTKEPVEPQPSQVPLQEDVDMSSGSSGHETNENCSTGRDSQGSDCDD 60
mesoper2          MNGYVNFSPSSTSPTKEPGEPQPSRAVLQEDVDMSSGSSGNETNENCSPGRDSQGSDCDD 60
cricetulus        MNGYVNFPPSSTSPTKEPGEPQPSQAALQEDVDMSSGSSGNETNENCSTGRDSQGSDCDD 60
mper2             MNGYVDFSPSPTSPTKEPGAPQPTQAVLQEDVDMSSGSSGN---ENCSTGRDSQGSDCDD 57
ratper2           MNGYVDFSPSPTSPTQEPGEPQPTQAVLQEDVDMSSGSSGN---ENCSTGRDSQGSDCDD 57
                  :**:.:* **  .**:*    **::. ************.   :*** ** ******:*

fpalper2          SGKELGMLVEPADTRESPGAF-HLMMTE---------PERNPSTSGCSSEQSAKADTHRE 110
ictidomysper2     NGKELGMLVETANTRESPRRA-CRSSPGHVPSEGCGTLTRTSVSWLHSSEQSAKADAHRE 116
hper2             SGKELGMLVEPPDARQSPD-TFSLMMAK---------SEHNPSTSGCSSDQSSKVDTHKE 110
mesoper2          NGKELRMLVEPSSAHPSPD-AFRLMMTE---------PEHNPSTSGCSSEQSAKADAHKE 110
cricetulus        NGKELRMLVEPSNTHPSPD-AFRLMMTE---------PEHNPSTSGCSSEQSAKADAHKE 110
mper2             NGKELRMLVESSNTHPSPDDAFRLMMTE---------AEHNPSTSGCSSEQSAKADAHKE 108
ratper2           SGKELRMLVESSNTHPSPDDTFRLMMTE---------AEHNPSTSGCSSEQSAKADAHKE 108
                  .**** ****  .:: **                    :.  :   **:**:*.*:*:*
```

```
fpalper2          LIRTLRELKVHLPADKQAKGKASTLATLKYALRSVKQVKANEEYFQLLMSSDSPPWGAGV 170
ictidomysper2     LVRTLRELKVHLPADKQAKGKASTLATLKYALRSVKQVKANEEYFQLLMSSDSQPCCAGV 176
hper2             LIKTLKELKVHLPADKKAKGKASTLATLKYALRSVKQVKANEEYYQLLMSSEGHPCGADV 170
mesoper2          LIRTLRELKVHLPADKKAKGKASTLATLKYALRSVKQVKANEEYYQLLMSSESQPCSVDV 170
cricetulus        LIRTLRELKVHLPADKKAKGKASTLATLKYALRSVKQVKANEEYYQLLMSSESQPCSVDV 170
mper2             LIRTLKELKVHLPADKKAKGKASTLATLKYALRSVKQVKANEEYYQLLMSSESQPCSVDV 168
ratper2           LIRTLRELKVHLPADKKAKGKASTLATLKYALRSVKQVKANEEYYQLLMSSESQPCSVDV 168
                  *::**:*********:**************************:******:. *  . *

fpalper2          PSYTVEQVEGVTSEYLLKNADMFAVAVSLVTGRILYISNQVTSIFHCKRDAFGDAKFVEF 230
ictidomysper2     PSYTVAQVEGATSEYVLKNADMFAVAVSLVTGKILYISDQVVSIFRCKREAFGDAQFVAF 236
hper2             PSYTVEEMESVTSEHIVKNADMFAVAVSLVSGKILYISDQVASIFHCKRDAFSDAKFVEF 230
mesoper2          PSYTMEQVEGITSEYIVKNADMFAVAVSLVSGKILYISNQVASIFHCKKDAFSDAKFVEF 230
cricetulus        PSYTMEQVEGITSEYIVKNADMFAVAVSLVSGKILYISNQVASIFHCKKDAFSDAKFVEF 230
mper2             PSYSMEQVEGITSEYIVKNADMFAVAVSLVSGKILYISNQVASIFHCKKDAFSDAKFVEF 228
ratper2           PSYTMEQVEGITSEYIVKNSDMFAVAVSLVSGKILYISNQVAPIFHCKKDAFSDAKFVEF 228
                  ***:: ::*. ***:::**:**********:*:*****:**. **:**::**.**:** *

fpalper2          LGPQDVSVFHSFTTPCKLPPWSACGGVDSFTQECMEEKSFFCRVGVGQNHENGIRYQPFR 290
ictidomysper2     LAPQDVSVFHSCTTPCKLPPWSMCGGVDSFTQECMEEKSFFCRVGVGENHENEIHYQPFR 296
hper2             LAPHDVGVFHSFTSPYKLPLWSMCSGADSFTQECMEEKSFFCRVSVRKSHENEIRYHPFR 290
mesoper2          LAPHDVSVFHSYTTPYKLPPWSVCSGLDSFTQECMEEKSFFCRVSVGKHHENEIRYQPFR 290
cricetulus        LAPHDVSVFHSCTTPYKLPPWSVCSGLDSFTQECMEEKSFFCRVSVGKHHENEIRYQPFR 290
mper2             LAPHDVSVFHSYTTPYKLPPWSVCSGLDSFTQECMEEKSFFCRVSVGKHHENEIRYQPFR 288
ratper2           LAPHDVSVFHSYTTPYKLPPWSVSSGLDSFTQECMEEKSFFCRVSVGKHHENEIRYQPFR 288
                  *.*:**.**** *:* *** ** ..* ****************.* : *** *:*:***

fpalper2          MTAYLAQVQEQPGAESQLCCLLLAERVHSGYEAPRIPPEKRIFTTTHTPNCLFQDVDERA 350
ictidomysper2     MTPYLAQVQERPGTESQLCCLLLAERVHSGYEAPRIPPEKRIFTTTHTPNCLFQDVDERA 356
hper2             MTPYLVKVRDQQGAESQLCCLLLAERVHSGYEAPRIPPEKRIFTTTHTPNCLFQDVDERA 350
mesoper2          MTPYLVKVQEQQGSESQLCCLLLAERVHSGYEAPRIPPEKRIFTTTHTPNCLFQDVDERA 350
cricetulus        MTPYLVKVQEQQGTESQLCCLLLAERVHSGYEAPRIPPEKRIFTTTHTPNCLFQDVDERA 350
mper2             MTPYLVKVQEQQGAESQLCCLLLAERVHSGYEAPRIPPEKRIFTTTHTPNCLFQAVDERA 348
ratper2           MTPYLVKVQEQKGAASQLCCLLLAERVHSGYEAPRIPPEKRIFTTTHTPNCLFQDVDERA 348
                  ** **.:*::: *: ************************************ *****

fpalper2          VPLLGHLPQDLIETPVLLQFHPSDRRLMLAIHKKILQSGGQPFDYSPMRFRTRNGEYITL 410
ictidomysper2     VPLLGYLPQDLMETPVLLQLHPSDRPLMLAIHKKILQSGGQPFDYSPMRFRARNGEYVTL 416
hper2             VPLLGYLPQDLIETPVLVQLHPSDRPLMLAIHKKILQSGGQPFDYSPIRFRARNGEYITL 410
mesoper2          VPLLGYLPQDLIETPVLVQLHPSDRPLMLAIHKKILQAGGQPFDYSPIRFRTRNGEYITL 410
cricetulus        VPLLGYLPQDLIETPVLVQLHPSDRPLMLAIHKKILQAGGQPFDYSPIRFRTRNGEYISL 410
mper2             VPLLGYLPQDLIETPVLVQLHPSDRPLMLAIHKKILQAGGQPFDYSPIRFRTRNGEYITL 408
ratper2           VPLLGYLPQDLIETPVLVQLHPSDRPLMLAIHKKILQASGQPFDYSPIRFRTRNGEYITL 408
                  *****:*****:*****:*:***** **********:.********:***:*****::*

fpalper2          DTSWSSFINPWSRKISFIIGRHRVRVGPLNEDVFAAPPCAEQKTLQPSVQELTEQIHRLL 470
ictidomysper2     DTSWSSFINPWSRKISFIIGRHRVRVGPLNEDVFAAPPCAEEKALPPNLQELAAQIHRLL 476
hper2             DTSWSSFINPWSRKISFIIGRHKVRVGPLNEDVFAAHPCTEEKALHPSIQELTEQIHRLL 470
mesoper2          DTSWSSFINPWSRKISFIIGRHKVRVGPLNEDVFAASPCPEEKTPHPSIQELTEQIHRLL 470
cricetulus        DTSWSSFINPWSRKISFIIGRHKVRVGPLNEDVFAASSCPEEKTPHPSVQELTEQIHRLL 470
mper2             DTSWSSFINPWSRKISFIIGRHKVRVGPLNEDVFAASPCPEEKTPHPSVQELTEQIHRLL 468
ratper2           DTSWSSFINPWSRKISFIIGRHKVRVGPLNEDVFAASPCPEEKTPHPSVQELTEQIHRLL 468
                  ********************:************  * *:*:  *.:***: ******
```

```
fpalper2          LQPVPHSGSSGYGSLGSNGSHEHLMSQTSSSDSNGHEDSRRWRSEICKNGN-KTKTKSHF 529
ictidomysper2     LQPVPHSGSSGYGSLGSNGSHEHLMSQTSSSDSNGHEESRQRRAEICKNGNDKTKTKTHF 536
hper2             LQPVPHSGSSGYGSLGSNGSHEHLMSQTSSSDSNGHEDSRRRRAEICKNGN-KTKNRSHY 529
mesoper2          MQPVPHSGSSGYGSLGSNGSHEHLMSQTSSSDSNGHEESHRKRSGIFKNGG-KIQTKSHF 529
cricetulus        MQPVPHSGSSGYGSLGSNGSHEHLMSQTSSSDSNGHEESHRKRPGIFKNGG-KIQTKSHF 529
mper2             MQPVPHSGSSGYGSLGSNGSHEHLMSQTSSSDSNGQEESHRRRSGIFKTSG-KIQTKSHV 527
ratper2           MQPVPHSGSSGYGSLGSNGSHEHLMSQTSSSDSNGQEESHWRRSGIFKTSG-KSQSKSHF 527
                  :**********************************:*:*:  *  * *..  * :.::*

fpalper2          PHASEEQKETSVTEMQSSPPAQAKAVPATEEDSLGASVPRTGVPEELACKDQPACSYQQI 589
ictidomysper2     SHESGEQKKKSITEMQSSPPAQGKDVPATEEDSLGASRPRASVPEELACKDQPTCSYQQI 596
hper2             SHESGEQKKKSVTEMQTNPPAEKKAVPAMEKDSLGVS-----FPEELACKNQPTCSYQQI 584
mesoper2          SHEAGEPKETSVAEMQSSPPAQVKAVTTIDRDSSGASLPKAVFPEELAYKNQPACSYQQI 589
cricetulus        SHESGEPKETSVAEMQSSPPAQVKAVTTVDRDSSGASLPKGSFPEELAYKNQPACSYQQI 589
mper2             SHESGGQKEASVAEMQSSPPAQVKAVTTIERDSSGASLPKASFPEELAYKNQPPCSYQQI 587
ratper2           SPESGGQKEASVAEMQSSPPAQVRSVTTMERDSSGASLPKASFPEELTYKSQPPCSYQQI 587
                     :   *: *::***:.***: : * : :.** *.*     .****: *.** ******

fpalper2          SCLDGVLRYLESCSEATTLKGKCEFPAHLPS------PKATTSPGLHTAGAASPSKVSSR 643
ictidomysper2     SCLDSVIRYLESCSEAATLKRKCASPANIPS------PKATASPGLHAGGAASPAKVSSR 650
hper2             SCLDSVIRYLESCNEAATLKRKCEFPANVPALRSSDKRKATVSPGPHAGEAEPPSRVNSR 644
mesoper2          SCLDSVIRYLESCSEAATLKRKCEFPANIPS------RKATVSPGLHSGEATRHSKVTSH 643
cricetulus        SCLDSVIRYLESCSEAATLKRKCEFPANIPS------RKATVSPGLHSGEAARHSKVTSH 643
mper2             SCLDSVIRYLESCSEAATLKRKCEFPANIPS------RKATVSPGLHSGEAARPSKVTSH 641
ratper2           SCLDSVIRYLESCNEAATLKRKCEFPANIPS------RKATVSPGLHSGEAARSSKVTSH 641
                  ****.*:******.**:*** **  **.:*:       ***.*** *:. *   ::*.*:

fpalper2          REVSARLSSLALPGKAESVVSLTSQCSYSSTIVHVGDKKPQPESEMAEDAASGPEALDGL 703
ictidomysper2     REVSARLSSLTLPGKAESVVSLTSQCSYSSTIVHVGDKKPQPESEMVEDAASGPESLDCL 710
hper2             TGVGTHLTSLALPGKAESVASLTSQCSYSSTIVHVGDKKPQPELEMVEDAASGPESLDCL 704
mesoper2          TEVSAHLSSLALPGKAESVVSLTSQCSYSSTIVHVGDKKPQPELETVEDVASGPESLDGA 703
cricetulus        TEVSAHLSSLTLPGKAESVVSLTSQCSYSSTIVHVGDKKPQPELETVEDVASGPESLDGT 703
mper2             TEVSAHLSSLTLPGKAESVVSLTSQCSYSSTIVHVGDKKPQPELETVEDMASGPESLDGA 701
ratper2           TEVSAHLSSLALPGKAESVVSLTSQCSYSSTIVHVGDKKPQPELETVEDVASGPESQDDA 701
                    *.::*:**:********.********************** * .** *****: *

fpalper2          P-----HSLSQEKGPLQKLGLTKEVLAAHTQREEQSFLQKFRGARRLGILQAHCHYYSRD 758
ictidomysper2     P-----HSLGQEKGPLQNLGLTKEVLAAHTQREEQSFLQKFRGARRLGVLPAHCHYYSRE 765
hper2             AGPALACGLSQEKEPFKKLGLTKEVLAAHTQKEEQSFLQKFKEIRKLSIFQSHCHYYLQE 764
mesoper2          A-----GGLSQEKGPLQKLGLTKEVLAAHTQREEQGFLQRFREVSRLGALQAHCQNYLQE 758
cricetulus        A-----GGLSQEKGPLQKLGLTKEVLAAHTQREEQGFLQRFRELSRLGALQARCHSYLQE 758
mper2             A-----GGLSQEKGPLQKLGLTKEVLAAHTQREEQGFLQRFREVSRLSALQAHCQNYLQE 756
ratper2           A-----GGLSQEKGSLQKLGLTKEVLAAHTQREEQGFLQRFREVSRLGALQAHCQNYLQE 756
                        .*.***  :::**************:***.***:*:   :*. : ::*: * ::

fpalper2          RSGGPPSERTAPGLRSASGIDSSWRKPGKGRKLKSRRAKPRGTPQSTGPGAPEPPRRPLI 818
ictidomysper2     RSGGPS---TAPGLRSTSGIDSSWKKTGKNRKLKSRRARPRGPSQSAGPGGLEPHRRPLA 822
hper2             RSKGQPSERTAPGLRNTSGIDSPWKKTGKNRKLKSKRVKPRDSSESTGSGGPVSARPPLV 824
mesoper2          RSRAQASERATPGPRNTSGIESSWKKTGKNRKLKSKRVKTQDSSESTGSGGPVSHRPPLV 818
cricetulus        RSRAQASERATPGLRNTSGIESSWKKTGKNRKLKSKRVKTQDSSESTGSGGPVSHRPPLV 818
mper2             RSRAQASDR---GLRNTSGLESSWKKTGKNRKLKSKRVKTRDSSESTGSGGPVSHRPPLM 813
ratper2           RSRAPASDR---GLRNASGIESSWKKTGKNRKLKSKRVKTRDSSESTGSGGPVSHRPPLV 813
                  ** .       * *.:**::* *:* ** *****:*.: :   :*:* *.    * **
```

```
fpalper2          SLSAMAWSPSDTSQSSCPAVAFPTSVP-TCPLPVFPAP------GPLVSPPADPHASLAM 871
ictidomysper2     SLSATAWSPSDTSQSSCPALAFPAAVP-TYPLPMFPAP------GTMVSPPAAPHTSFAV 875
hper2             GLNATAWSPSDTSQSSCPAVPFPAPVPAAYSLPVFPAP------GTVAAPPAPPHASFTV 878
mesoper2          GLNATAWSPSDTSQSSCPSAPFPASVP-AYPLPVFQAPGIVSTPGTVVAPPAAAHSSFTM 877
cricetulus        GLNATAWSPSDTSQSSCPSAPFPASVP-AYPLPVFQAPGIVSTPGTVGAPPAPAHPGFTM 877
mper2             GLNATAWSPSDTSQSSCPSAPFPTAVP-AYPLPVFQAPGIVSTPGTVVAPPAATHTGFTM 872
ratper2           GLNATAWSPSDTSQSSCPSAPFPAPVP-AYPLPVFPAPGIVSTPGTVVAPPAAAHTGFTM 872
                  .*.* ************:  **: ** :  **:* **      * : :***  * .:::

fpalper2          PTLPVDTQPEFAVQPLPFAAPLAPVMALVLPTCPFPPVAPGLPQAFFPGQPHFPGHPTLP 931
ictidomysper2     PALPVGPQHEFAVQPLPFSAPVAPLVALMLPTCPFPPVAPNLPQAFFPG------HPRLP 929
hper2             PAVPVDLQHQFAVQPPPFPAPLAPVMAFMLPSYSFPSGTPNLPQAFFPSQPQFPSHPTLT 938
mesoper2          PVVPVGTHPEFAVQPLPFAAPLAPVMAFMLPSYPFPPATPNLPQAFFPSQPSFPAHPTLA 937
cricetulus        PVVPMGTHPEFAVQPLTFAAPLTPVMAFMLPSYPFPPATPNLPQAFFPSQPPFPAHPTLA 937
mper2             PVVPMGTQPEFAVQPLPFAAPLAPVMAFMLPSYPFPPATPNLPQAFLPSQPHFPAHPTLA 932
ratper2           PVVPMGTQPEFAVQPLPFAAPLAPVMAFMLPSYPFPPATPNLPQAFFPSQPHFPAHPTLA 932
                  *.:*:  : :*****  * **::*::*::**:  **  :* *****:*.     ** *

fpalper2          SEMRPAPQPELPHQTSLPGQPCACPATPPSAMVATGRASPPLFQSRGSSPLQLNLLQMEE 991
ictidomysper2     SGVTPGSQPEAPQQTPLPRRPCPRPATPPSATVAAGRASPPLFQSRGSSPLQLNLLQMEE 989
hper2             SEMASASQPEFPSRTSIPRQPCACPATRATPPSAMGRASPPLFQSRSSSPLQLNLLQLEE 998
mesoper2          SEITPASQAEFPSRTSVLRQPCTCPVTPPAGTAASGRASPPLFQSRGSSPLQLNLLQLEE 997
cricetulus        SEVAPASQAEFPSRTSVLRQSCTCPATPPAGTVASGRASPPLFQSRGSSPLQLNLLQLEE 997

mper2             SEITPASQAEFPSRTSTLRQPCACPVTPPAGTVALGRASPPLFQSRGSSPLQLNLLQLEE 992
ratper2           SEITPASQAEFPSRTSMLRQPCACPVTPPAGTVALGRASPPLFQSRGSSPLQLNLLQLEE 992
                  * :  . * * * :*    : *  *.*  :   * **********.**********:**

fpalper2          PPEGSAGAAAPLGTPGAAAARPDCTPGASRDPQPKAPPTRDEPLDAQESDALSVSSDLLN 1051
ictidomysper2     PPEGSAAATATLGHPGTAAARPDCTPGAPRDQQAKAPPTREEPSDAQESDALSVSSDLLN 1049
hper2             APEGGTGAMGTTGATETAAVGADCKPGTSRDQQPKAPLTRDEPSDTQNSDALSTSSGLLN 1058
mesoper2          APESSIGAPGTLGTTGTAASGLDCATSTSRDRQPMASPTCNEPSDTQNSDAISTSSDLLN 1057
cricetulus        APESSIGAPGTLGTTGTAASGLDCTAGTSRDRQPKAPPTCNEPSDTQNSDAISTSSDLLN 1057
mper2             APEGSTGAAGTLGTTGTAASGLDCTSGTSRDRQPKAPPTCNEPSDTQNSDAISTSSDLLN 1052
ratper2           APESSTGAAGTLGTTGTAASGLDCTSGASRDRQPKAPPTCSEPSDTQNSDAISTSSDLLN 1052
                   **.. .* .  *   :**   **  .: ** *  *  * .** *:*:***:*.** ***

fpalper2          LLLSEDVCSAAGSALSGSGASAASDS-----LGCGASQSGAGSSDTSRTSKYFGSIDSSE 1106
ictidomysper2     LLLSEDLCSATGSALSGSGASATSGSLGSSSPGCSASHSGAGSSDTSRTSKYFGSIDSSE 1109
hper2             LLLNEDLCSASGSA--------ASESLGSGSLGCDASPSGAGSSDTSHTSKYFGSIDSSE 1110
mesoper2          LLLGEDLCSATGSALSRSGASATSDSLGSSSLGCDVSRSGAGSSDTSHTSKYFGSIDSSE 1117
cricetulus        LLLGEDLCSATGSALSRSGASATSDSLGSSSLGCDASRSGAGSSDTSHTSKYFGSIDSSE 1117
mper2             LLLGEDLCSATGSALSRSGASATSDSLGSSSLGFGTSQSGAGSSDTSHTSKYFGSIDSSE 1112
ratper2           LLLGEDLCSATGSALSRSGASATSDSLGSSSLGCDTSRSGAGSSDTSHTSKYFGSIDSSE 1112
                  *** **:***:***       :* *     *  .* *********:***********

fpalper2          NNHRAEMNRDKEQSEHFIKYVLQDPIWLLMADADDSVMMTYQLPSRDLESVLKEDREKLK 1166
ictidomysper2     NHHKAEMDRSKGQSEHFIKYVLQDPIWLLMADTDDSIMMTYQLPSRDLESVLKEDREKLK 1169
hper2             NNHKAKMNTGMEESEHFIKCVLQDPIWLLMADADSSVMMTYQLPSRNLEAVLKEDREKLK 1170
mesoper2          NNHKAKMITDTEESEQFIKYVLQDPIWLLMANTDDSIMMTYQLPSRDLQSVLKEDREKLK 1177
cricetulus        NNHKAKVTTDTEESEQFIKYVLQDPIWLLMANTDDSIMMTYQLPSRDLQSVLKEDREKLK 1177
mper2             NNHKAKMIPDTEESEQFIKYVLQDPIWLLMANTDDSIMMTYQLPSRDLQAVLKEDQEKLK 1172
ratper2           NNHKAKMITDTEESEQFIKYVLQDPIWLLMANTDDNIMMTYQLPSRDLQAVLKEDQEKLK 1172
                  *.*:*::     :**:*** **********::*..:*********:*::*****:****
```

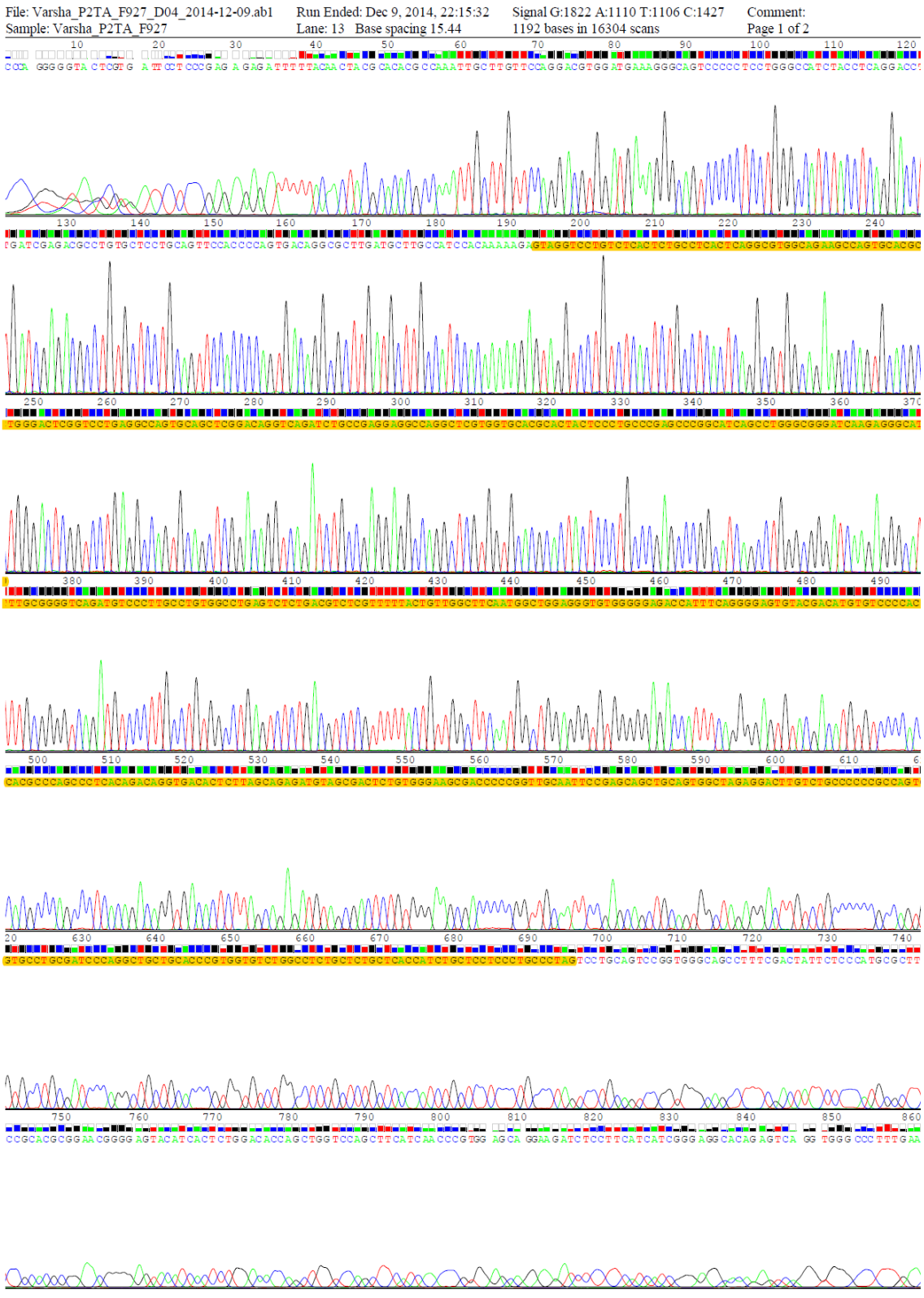
```
fpalper2          LLQRSQPWFTEGQRQELREVHPWVQTGGLPAAIDVAECVYCDSEKKGSVCVPFEGDIPSL 1226
ictidomysper2     LLQRSQPWFTEGQRQELREVHPWVQTGGLPAAINVAECVYCESEEKGSICVPYEGDIPSL 1229
hper2             LLQKLQPRFTESQKQELREVHQWMQTGGLPAAIDVAECVYCENKEKGNICIPYEEDIPSL 1230
mesoper2          LLQRSQPRFTEGQRRELRDVHPWVQTGGLPTAIDVTECIYCESEEKGNICLPYEEDSPPL 1237
cricetulus        LLQRSQPQFTEGQRRELREVHPWVQTGGLPAAIDVTECIYCESEEKGNICLPYEEDSPPL 1237
mper2             LLQRSQPRFTEGQRRELREVHPWVHTGGLPTAIDVTGCVYCESEEKGNICLPYEEDSPSP 1232
ratper2           LLQRSQPHFTEGQRRELREVHPWVHTGGLPTAIDVTGCVYCESEEKGNLCLPYEEDSPSL 1232
                  ***: ** ***.*::***:** *::*****:**:*: *:**:.::**.:*:*:* * *

fpalper2          GLSDMSDTKEEESGHPPRHREKERT    1251
ictidomysper2     GLSDASDTKEEESGQAPRPRAGEQT    1254
hper2             GLSEVSDTKEDENGSPLNHRIEEQT    1255
mesoper2          GLADTSETKEEESGQLKSPRKETQT    1262
cricetulus        GLTDTSETKEEESGQLSSPRKEAQT    1262
mper2             GLCDTSEAKEEEGEQLTGPRIEAQT    1257
ratper2           GLCDTSEAKEEESGQLANPRKEAQT    1257
                  ** : *::**:*       *   :*
```

F. palmarum Per2 variant, *F. pal Per2v*

C25. Sequencing *F. palmarum* CDS in TA XL vector (colony b). Primer F927

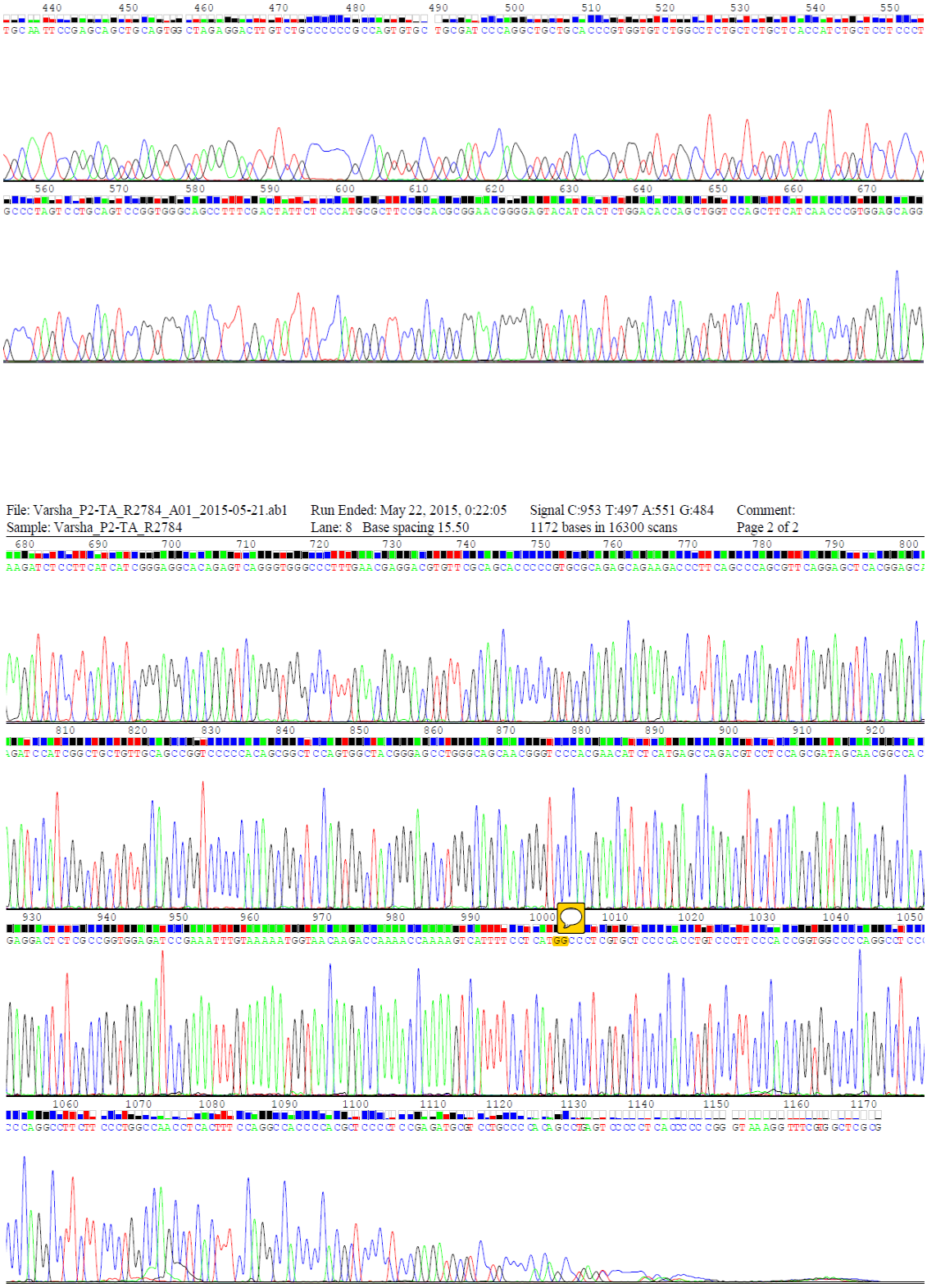
The region which is not present in the basic transcript is highlighted.

(only the chromatograms that indicates difference from basic transcript is included)



C 26. Sequencing *F. palmarum* CDS in TA XL vector (colony b). Primer R2784

The chromatogram is Reverse complimented, nucleotides between G/G which is highlighted is missing in this variant



D4. Complete nucleotide sequence of *F. palmarum Per2v.*

Bases indicated in red is an insertion, the G/G highlighted indicates missing bases between their junction. 'TGA' is a PTC as a consequence of the insertion. TAA is the normal stop codon as the basic transcript.

GTGAATGGATATGCGGAATTTCCCCCCAGCCCCAGCAACCCCACCAAGGA
GCCCATGGACCCCCAGCCCAGTCGGGCCTCACTTCAGGAGGACGTGGACAT
GAGCAGTGGCTCCAGTGGAAATGAAACCAATGACAACTGCTCCGTGGGGC
GGGGCTCACAGGGCAGCGACTGCGACGACAGCGGGAAGGAGCTGGGGATG
CTGGTGGAGCCCGCCGACACCCGGGAGAGCCCCGGGGCCTTTCACCTCATG
ATGACAGAGCCTGAGCGCAACCCGTCCACCAGCGGCTGCAGCAGCGAGCA
GTCTGCCAAAGCAGACACGCACAGGGAGCTGATAAGAACGCTGAGGGAGC
TGAAGGTCCACCTCCCTGCAGACAAGCAGGCCAAGGGCAAGGCCAGCACC
CTGGCGACCCTGAAGTACGCCCTGAGGAGCGTGAAGCAGGTGAAAGCTAA
TGAGGAGTATTTCCAGCTGCTCATGTCCAGCGACAGCCCGCCCTGGGGTGC
GGGTGTGCCCTCCTACACGGTGGAGCAGGTGGAGGGCGTCACCTCCGAGTA
CCTCCTGAAGAACGCGGATATGTTTGCTGTGGCCGTGTCCCTGGTCACTGGG
AGGATCCTCTATATCTCTAACCAAGTCACGTCCATATTCCATTGTAAGAGAG
ATGCCTTTGGTGACGCAAAGTTTGTGGAGTTCCAGGGCCCGCAGGACGTCA
GCGTGTTCCACAGCTTCACCACCCCTTGCAAGCTCCCGCCCTGGAGCGCGTG
CGGCGGAGTAGATTCTTTTACTCAAGAGTGCATGGAGGAGAAATCTTTCTT
CTGCCGTGTGGGGGTCGGGCAGAACCACGAGAATGGGATCCGCTACCAGCC
GTTCCGCATGACAGCCTACCTGGCCCAGGTGCAGGAGCAGCCGGGCGCTGA
GAGCCAGCTCTGCTGCCTGCTGCTGGCAGAGAGGGTGCACTCTGGCTATGA
AGCTCCTAGAATTCCTCCCGAGAAGAGAATTTTTACAACTACGCACACGCC
AAATTGCTTGTTCCAGGACGTGGATGAAAGGGCAGTCCCCCTCCTGGGCCA
TCTACCTCAGGACCTGATCGAGACGCCTGTGCTCCTGCAGTTCCACCCCAGT
GACAGGCGCTTGATGCTTGCCATCCACAAAAAAGAGTAGGTCCTGTCTCACT
CTGCCTCACTCAGGCGTGGCAGAAGCCAGTGCACGCTGGGACTCGGTCCTG
AGGCCAGTGCAGCTCGGACAGGTCAGATCTGCCGAGGAGGCCAGGCTCGT
GGTGCACGCACTACTCCCTGCCCGAGCCCGGCATCAGCCTGGGCGGGATCA
AGAGGGCATTTGCGGGGTCAGATGTCCCTTGCCTGTGGCCTGAGTCTCTG
ACGTCTCGTTTTTACTGTTGGCTTCAATGGCTGGAGGGTGTGGGGGAGACC
ATTTCAGGGGAGTGTACGACATGTGTCCCCACCACGCCCAGCCCTCACAGA
CAGGTGACACTCTTAGCAGAGATGTAGCGACTCTGTGGGAAGCGACCCCCG
GTTGCAATTCCGAGCAGCTGCAGTGGCTAGAGGACTTGTCTGCCCCCCGCC
AGTGTGCCTGCGATCCCAGGCTGCTGCACCCGTGGTGTCTGGCCTCTGCTCT
GCTCACCATCTGCTCCTCCCTGCCCTAGTCCTGCAGTCCGGTGGGCAGCCTT
TCGACTATTCTCCCATGCGCTTCCGCACGCGGAACGGGGAGTACATCACTCT
GGACACCAGCTGGTCCAGCTTCATCAACCCGTGGAGCAGGAAGATCTCCTT
CATCATCGGGAGGCACAGAGTCAGGGTGGGCCCTTTGAACGAGGACGTGTT
CGCAGCACCCCCGTGCGCAGAGCAGAAGACCCTTCAGCCCAGCGTTCAGGA
GCTCACGGAGCAGATCCATCGGCTGCTGTTGCAGCCGGTCCCCCACAGCGG

CTCCAGTGGCTACGGGAGCCTGGGCAGCAACGGGTCCCACGAACATCTCAT
GAGCCAGACGTCCTCCAGCGATAGCAACGGCCACGAGGACTCTCGCCGGTG
GAGATCCGAAATTTGTAAAAATGGTAACAAGACCAAAACCAAAAGTCATT*T*
TCCTCATGGCCCTCGTGCTCCCCACCTGTCCCTTCCCACCGGTGGCCCCAGG
CCTCCCCCAGGCCTTCTTCCCTGGCCAGCCTCACTTTCCAGGCCACCCCACG
CTCCCCTCCGAGATGCGTCCTGCCCCACAGCCTGAGCTCCCCCATCAGACCT
CACTCCCAGGACAGCCCTGCGCTTGCCCAGCCACCCCACCGTCAGCCATGG
TGGCCACGGGCAGGGCCTCCCCGCCGCTCTTCCAGTCCCGAGGCAGCTCGC
CCCTGCAGCTCAACCTGCTGCAGATGGAGGAGCCCCCCGAGGGCAGCGCCG
GGGCTGCGGCACCCCTGGGCACTCCTGGAGCAGCAGCCGCCAGGCCGGACT
GCACCCCTGGCGCTTCTCGGGACCCGCAGCCGAAGGCACCCCCAACGCGAG
ACGAGCCCCTAGATGCACAGGAGAGTGACGCCCTGTCTGTGTCCAGCGACC
TGCTCAACCTCCTGCTGAGTGAGGACGTCTGCTCTGCCGCGGGGTCAGCCCT
CTCTGGGAGTGGGGCCTCTGCCGCCTCGGATTCTCTGGGCTGCGGTGCATCC
CAAAGTGGGGCAGGCAGCAGTGACACCAGTCGTACCAGCAAATACTTTGG
AAGCATCGACTCTTCAGAGAATAATCACAGAGCAGAAATGAACAGGGACA
AGGAGCAGAGTGAACACTTCATCAAGTACGTCCTTCAGGACCCCATCTGGC
TGCTGATGGCGGACGCGGACGACAGCGTCATGATGACCTACCAGCTGCCTT
CCCGGGATCTGGAGTCAGTCTTGAAGGAGGACAGGGAGAAGCTGAAGCTA
CTGCAGAGGTCCCAGCCCTGGTTCACTGAGGGCCAGAGGCAGGAGCTGCGT
GAGGTCCACCCATGGGTCCAGACAGGTGGCCTGCCCGCGGCCATTGATGTG
GCGGAATGTGTTTACTGTGACAGCGAAGAGAAAGGCAGTGTTTGCGTGCCA
TTTGAGGGAGACATTCCTTCCCTGGGACTCAGCGACATGTCCGACACCAAA
GAAGAGGAAAGCGGACACCCCCCGAGGCACAGGGAGAAGGAGCGGACG
TAACGCCCCGACAGCCAGGGATCCGCGTCAGATGGTGCTTGGAAGACAG
AGGCTCCTCACGCTTGTTTCTCGTGAAATGGGTATAGATGCGTCCATTGCTC
TTTGTCTTAGGGGAAAACCCCAGTTTTCTGAAGGGGTGATTTAAAACTGGA
GGGTAGGAAGGGTTTAGGAAGAAATACGTTTTTGTATTTAAAATGTAAACG
TGGAGTTTGGGCACGGAGCAGAGATTTTGTTTGTTACTGAACTTGAGGAAG
GCTGAGACCCGTCTTTGGAATTATCAGGGAAGTTGCCTCAGTGCCCTTGGA
AGTTCTTCAACGAAGATGAGACCTCTGTCCAGTGCACCAGCATCCGCTGGC
CCCCTGTGGGTTGACCCAAGACGTCACAGCAGGCTGTGACAAATAGCCAGG
AAGCCAGGAGTGGCTCTGAGCTGAGGGCTCCGGTTTTCTGGAGCTTTTCCA
GGGTGTCTTTAGATTGGATCAAATGTTGTCCTCTCAGGGGCTTGTCCAAGTG
GTCACAGGTCCACCGGATGTTTGCTCATATACCGCATGGAGGTAACGGCCT
CTCTTCACTTTCTTTCAAAGGCTTCACATTAGGTGGGGTGCAGCATGATTCA
ACAGCTCTGGCGTCCAGTACTGTAGTGGTAGACATTCTGGTGTGACGTGGC
TTCTGTTTGTAGTGAATAGAGTTGAAATGCTCACTTCGGACTAACAGATCCT
AGTGGTTTATGTGACCAGCCTCACCAGATTGATGCCCACCCAGACCCTCTGT
ATATTTCCCCATAATTCTTATATGCCTTTTAGACCAGATGGTCAATAGTAGA
GAGAAATATGTTACCAGCCATGTCCTCACATCTTCATGGTGACCTCCCTCTG
CAAGCCAGGCCTTGACCACTACGCAGGGGACCCAGTCGTGTTTGTCTTTTG
GTTGGAGCCTCTCAGACATCAGCTCTCACCCAGC

A5. CLUSTAL alignment of *F. palmarum Per2* and *Per2v*.

Alignment is done using the 1st 1153 bases followed separately by the remaining bases of transcript

3a is the basic transcript, 3b is the colony from where Per2v was found

```
                *         20         *         40
3a  :  GTGAATGGATATGCGGAATTTCCCCCCAGCCCCAGCAACCCCACC  :    45
3b  :  GTGAATGGATATGCGGAATTTCCCCCCAGCCCCAGCAACCCCACC  :    45
       GTGAATGGATATGCGGAATTTCCCCCCAGCCCCAGCAACCCCACC

           *         60         *         80         *
3a  :  AAGGAGCCCATGGACCCCCAGCCCAGTCGGGCCTCACTTCAGGAG  :    90
3b  :  AAGGAGCCCATGGACCCCCAGCCCAGTCGGGCCTCACTTCAGGAG  :    90
       AAGGAGCCCATGGACCCCCAGCCCAGTCGGGCCTCACTTCAGGAG

                100         *        120         *
3a  :  GACGTGGACATGAGCAGTGGCTCCAGTGGAAATGAAACCAATGAC  :   135
3b  :  GACGTGGACATGAGCAGTGGCTCCAGTGGAAATGAAACCAATGAC  :   135
       GACGTGGACATGAGCAGTGGCTCCAGTGGAAATGAAACCAATGAC

              140         *        160         *        180
3a  :  AACTGCTCCGTGGGGCGGGGCTCACAGGGCAGCGACTGCGACGAC  :   180
3b  :  AACTGCTCCGTGGGGCGGGGCTCACAGGGCAGCGACTGCGACGAC  :   180
       AACTGCTCCGTGGGGCGGGGCTCACAGGGCAGCGACTGCGACGAC

                *        200         *        220
3a  :  AGCGGGAAGGAGCTGGGGATGCTGGTGGAGCCCGCCGACACCCGG  :   225
3b  :  AGCGGGAAGGAGCTGGGGATGCTGGTGGAGCCCGCCGACACCCGG  :   225
       AGCGGGAAGGAGCTGGGGATGCTGGTGGAGCCCGCCGACACCCGG

           *        240         *        260         *
3a  :  GAGAGCCCCGGGGCCTTTCACCTCATGATGACAGAGCCTGAGCGC  :   270
3b  :  GAGAGCCCCGGGGCCTTTCACCTCATGATGACAGAGCCTGAGCGC  :   270
       GAGAGCCCCGGGGCCTTTCACCTCATGATGACAGAGCCTGAGCGC

                280         *        300         *
3a  :  AACCCGTCCACCAGCGGCTGCAGCAGCGAGCAGTCTGCCAAAGCA  :   315
3b  :  AACCCGTCCACCAGCGGCTGCAGCAGCGAGCAGTCTGCCAAAGCA  :   315
       AACCCGTCCACCAGCGGCTGCAGCAGCGAGCAGTCTGCCAAAGCA

              320         *        340         *        360
3a  :  GACACGCACAGGGAGCTGATAAGAACGCTGAGGGAGCTGAAGGTC  :   360
3b  :  GACACGCACAGGGAGCTGATAAGAACGCTGAGGGAGCTGAAGGTC  :   360
       GACACGCACAGGGAGCTGATAAGAACGCTGAGGGAGCTGAAGGTC

                *        380         *        400
3a  :  CACCTCCCTGCAGACAAGCAGGCCAAGGGCAAGGCCAGCACCCTG  :   405
3b  :  CACCTCCCTGCAGACAAGCAGGCCAAGGGCAAGGCCAGCACCCTG  :   405
       CACCTCCCTGCAGACAAGCAGGCCAAGGGCAAGGCCAGCACCCTG

           *        420         *        440         *
3a  :  GCGACCCTGAAGTACGCCCTGAGGAGCGTGAAGCAGGTGAAAGCT  :   450
3b  :  GCGACCCTGAAGTACGCCCTGAGGAGCGTGAAGCAGGTGAAAGCT  :   450
       GCGACCCTGAAGTACGCCCTGAGGAGCGTGAAGCAGGTGAAAGCT
```

```
          460         *       480         *
3a : AATGAGGAGTATTTCCAGCTGCTCATGTCCAGCGACAGCCCGCCC : 495
3b : AATGAGGAGTATTTCCAGCTGCTCATGTCCAGCGACAGCCCGCCC : 495
     AATGAGGAGTATTTCCAGCTGCTCATGTCCAGCGACAGCCCGCCC

       500         *       520         *       540
3a : TGGGGTGCGGGTGTGCCCTCCTACACGGTGGAGCAGGTGGAGGGC : 540
3b : TGGGGTGCGGGTGTGCCCTCCTACACGGTGGAGCAGGTGGAGGGC : 540
     TGGGGTGCGGGTGTGCCCTCCTACACGGTGGAGCAGGTGGAGGGC

             *       560         *       580
3a : GTCACCTCCGAGTACCTCCTGAAGAACGCGGATATGTTTGCTGTG : 585
3b : GTCACCTCCGAGTACCTCCTGAAGAACGCGGATATGTTTGCTGTG : 585
     GTCACCTCCGAGTACCTCCTGAAGAACGCGGATATGTTTGCTGTG

         *       600         *       620         *
3a : GCCGTGTCCCTGGTCACTGGGAGGATCCTCTATATCTCTAACCAA : 630
3b : GCCGTGTCCCTGGTCACTGGGAGGATCCTCTATATCTCTAACCAA : 630
     GCCGTGTCCCTGGTCACTGGGAGGATCCTCTATATCTCTAACCAA

          640         *       660         *
3a : GTCACGTCCATATTCCATTGTAAGAGAGATGCCTTTGGTGACGCA : 675
3b : GTCACGTCCATATTCCATTGTAAGAGAGATGCCTTTGGTGACGCA : 675
     GTCACGTCCATATTCCATTGTAAGAGAGATGCCTTTGGTGACGCA

       680         *       700         *       720
3a : AAGTTTGTGGAGTTCCTGGGCCCGCAGGACGTCAGCGTGTTCCAC : 720
3b : AAGTTTGTGGAGTTCCAGGGCCCGCAGGACGTCAGCGTGTTCCAC : 720
     AAGTTTGTGGAGTTCCAGGGCCCGCAGGACGTCAGCGTGTTCCAC

             *       740         *       760
3a : AGCTTCACGACCCCTTGCAAGCTCCCGCCCTGGAGCGCGTGCGGC : 765
3b : AGCTTCACCACCCCTTGCAAGCTCCCGCCCTGGAGCGCGTGCGGC : 765
     AGCTTCACCACCCCTTGCAAGCTCCCGCCCTGGAGCGCGTGCGGC

         *       780         *       800         *
3a : GGAGTAGATTCTTTTACTCAAGAGTGCATGGAGGAGAAATCTTTT : 810
3b : GGAGTAGATTCTTTTACTCAAGAGTGCATGGAGGAGAAATCTTTC : 810
     GGAGTAGATTCTTTTACTCAAGAGTGCATGGAGGAGAAATCTTTC

          820         *       840         *
3a : TTCTGCCGTGTGGGGGTCGGGCAGAACCACGAGAATGGGATCCGC : 855
3b : TTCTGCCGTGTGGGGGTCGGGCAGAACCACGAGAATGGGATCCGC : 855
     TTCTGCCGTGTGGGGGTCGGGCAGAACCACGAGAATGGGATCCGC

       860         *       880         *       900
3a : TACCAGCCGTTCCGCATGACAGCCTACCTGGCCCAGGTGCAGGAG : 900
3b : TACCAGCCGTTCCGCATGACAGCCTACCTGGCCCAGGTGCAGGAG : 900
     TACCAGCCGTTCCGCATGACAGCCTACCTGGCCCAGGTGCAGGAG
```

```
                *       920         *       940
3a  : CAGCCGGGCGCTGAGAGCCAGCTCTGCTGCCTGCTGCTGGCAGAG :  945
3b  : CAGCCGGGCGCTGAGAGCCAGCTCTGCTGCCTGCTGCTGGCAGAG :  945
      CAGCCGGGCGCTGAGAGCCAGCTCTGCTGCCTGCTGCTGGCAGAG

            *       960         *       980         *
3a  : AGGGTGCACTCTGGCTATGAAGCTCCTAGAATTCCTCCCGAGAAG :  990
3b  : AGGGTGCACTCTGGCTATGAAGCTCCTAGAATTCCTCCCGAGAAG :  990
      AGGGTGCACTCTGGCTATGAAGCTCCTAGAATTCCTCCCGAGAAG

          1000          *      1020          *
3a  : AGAATTTTTACAACTACGCACACGCCAAATTGCTTGTTCCAGGAC : 1035
3b  : AGAATTTTTACAACTACGCACACGCCAAATTGCTTGTTCCAGGAC : 1035
      AGAATTTTTACAACTACGCACACGCCAAATTGCTTGTTCCAGGAC

       1040          *      1060          *      1080
3a  : GTGGATGAAAGGGCAGTCCCCCTCCTGGGCCATCTACCTCAGGAC : 1080
3b  : GTGGATGAAAGGGCAGTCCCCCTCCTGGGCCATCTACCTCAGGAC : 1080
      GTGGATGAAAGGGCAGTCCCCCTCCTGGGCCATCTACCTCAGGAC

                *      1100          *      1120
3a  : CTGATCGAGACGCCTGTGCTCCTGCAGTTCCACCCCAGTGACAGG : 1125
3b  : CTGATCGAGACGCCTGTGCTCCTGCAGTTCCACCCCAGTGACAGG : 1125
      CTGATCGAGACGCCTGTGCTCCTGCAGTTCCACCCCAGTGACAGG

            *      1140          *      1160          *
3a  : CGCTTGATGCTTGCCATCCACAAAAAGA------------------ : 1153
3b  : CGCTTGATGCTTGCCATCCACAAAAAGAGTAGGTCCTGTCTCACT : 1170
      CGCTTGATGCTTGCCATCCACAAAAAGAGTAGGTCCTGTCTCACT

          1180          *      1200          *
3a  : ---------------------------------------------- :    -
3b  : CTGCCTCACTCAGGCGTGGCAGAAGCCAGTGCACGCTGGGACTCG : 1215
      CTGCCTCACTCAGGCGTGGCAGAAGCCAGTGCACGCTGGGACTCG

       1220          *      1240          *      1260
3a  : ---------------------------------------------- :    -
3b  : GTCCTGAGGCCAGTGCAGCTCGGACAGGTCAGATCTGCCGAGGAG : 1260
      GTCCTGAGGCCAGTGCAGCTCGGACAGGTCAGATCTGCCGAGGAG

                *      1280          *      1300
3a  : ---------------------------------------------- :    -
3b  : GCCAGGCTCGTGGTGCACGCACTACTCCCTGCCCGAGCCCGGCAT : 1305
      GCCAGGCTCGTGGTGCACGCACTACTCCCTGCCCGAGCCCGGCAT

            *      1320          *      1340          *
3a  : ---------------------------------------------- :    -
3b  : CAGCCTGGGCGGGATCAAGAGGGCATTTGCGGGGTCAGATGTCCC : 1350
      CAGCCTGGGCGGGATCAAGAGGGCATTTGCGGGGTCAGATGTCCC
```

```
          1360         *      1380         *
3a : ------------------------------------------- :    -
3b : TTGCCTGTGGCCTGAGTCTCTGACGTCTCGTTTTTACTGTTGGCT : 1395
     TTGCCTGTGGCCTGAGTCTCTGACGTCTCGTTTTTACTGTTGGCT

      1400         *      1420         *      1440
3a : ------------------------------------------- :    -
3b : TCAATGGCTGGAGGGTGTGGGGGAGACCATTTCAGGGGAGTGTAC : 1440
     TCAATGGCTGGAGGGTGTGGGGGAGACCATTTCAGGGGAGTGTAC

             *      1460         *      1480
3a : ------------------------------------------- :    -
3b : GACATGTGTCCCCACCACGCCCAGCCCTCACAGACAGGTGACACT : 1485
     GACATGTGTCCCCACCACGCCCAGCCCTCACAGACAGGTGACACT

         *      1500         *      1520         *
3a : ------------------------------------------- :    -
3b : CTTAGCAGAGATGTAGCGACTCTGTGGGAAGCGACCCCCGGTTGC : 1530
     CTTAGCAGAGATGTAGCGACTCTGTGGGAAGCGACCCCCGGTTGC

          1540         *      1560         *
3a : ------------------------------------------- :    -
3b : AATTCCGAGCAGCTGCAGTGGCTAGAGGACTTGTCTGCCCCCCGC : 1575
     AATTCCGAGCAGCTGCAGTGGCTAGAGGACTTGTCTGCCCCCCGC

      1580         *      1600         *      1620
3a : ------------------------------------------- :    -
3b : CAGTGTGCCTGCGATCCCAGGCTGCTGCACCCGTGGTGTCTGGCC : 1620
     CAGTGTGCCTGCGATCCCAGGCTGCTGCACCCGTGGTGTCTGGCC

             *      1640         *
3a : ---------------------------------- :    -
3b : TCTGCTCTGCTCACCATCTGCTCCTCCCTGCCCTAG : 1656
     TCTGCTCTGCTCACCATCTGCTCCTCCCTGCCCTAG
```

Second part of the Alignment

```
                *         20          *         40
3b  :  TCCTGCAGTCCGGTGGGCAGCCTTTCGACTATTCTCCCATGCGCT  :    45
3a  :  TCCTGCAGTCCGGTGGGCAGCCTTTCGACTATTCTCCCATGCGCT  :    45
       TCCTGCAGTCCGGTGGGCAGCCTTTCGACTATTCTCCCATGCGCT

            *         60          *         80          *
3b  :  TCCGCACGCGGAACGGGGAGTACATCACTCTGGACACCAGCTGGT  :    90
3a  :  TCCGCACGCGGAACGGGGAGTACATCACTCTGGACACCAGCTGGT  :    90
       TCCGCACGCGGAACGGGGAGTACATCACTCTGGACACCAGCTGGT

              100         *         120         *
3b  :  CCAGCTTCATCAACCCGTGGAGCAGGAAGATCTCCTTCATCATCG  :   135
3a  :  CCAGCTTCATCAACCCGTGGAGCAGGAAGATCTCCTTCATCATCG  :   135
       CCAGCTTCATCAACCCGTGGAGCAGGAAGATCTCCTTCATCATCG

         140         *         160         *         180
3b  :  GGAGGCACAGAGTCAGGGTGGGCCCTTTGAACGAGGACGTGTTCG  :   180
3a  :  GGAGGCACAGAGTCAGGGTGGGCCCTTTGAACGAGGACGTGTTCG  :   180
       GGAGGCACAGAGTCAGGGTGGGCCCTTTGAACGAGGACGTGTTCG

                *         200         *         220
3b  :  CAGCACCCCCGTGCGCAGAGCAGAAGACCCTTCAGCCCAGCGTTC  :   225
3a  :  CAGCACCCCCGTGCGCAGAGCAGAAGACCCTTCAGCCCAGCGTTC  :   225
       CAGCACCCCCGTGCGCAGAGCAGAAGACCCTTCAGCCCAGCGTTC

            *         240         *         260         *
3b  :  AGGAGCTCACGGAGCAGATCCATCGGCTGCTGTTGCAGCCGGTCC  :   270
3a  :  AGGAGCTCACGGAGCAGATCCATCGGCTGCTGTTGCAGCCGGTCC  :   270
       AGGAGCTCACGGAGCAGATCCATCGGCTGCTGTTGCAGCCGGTCC

              280         *         300         *
3b  :  CCCACAGCGGCTCCAGTGGCTACGGGAGCCTGGGCAGCAACGGGT  :   315
3a  :  CCCACAGCGGCTCCAGTGGCTACGGGAGCCTGGGCAGCAACGGGT  :   315
       CCCACAGCGGCTCCAGTGGCTACGGGAGCCTGGGCAGCAACGGGT

         320         *         340         *         360
3b  :  CCCACGAACATCTCATGAGCCAGACGTCCTCCAGCGATAGCAACG  :   360
3a  :  CCCACGAACATCTCATGAGCCAGACGTCCTCCAGCGATAGCAACG  :   360
       CCCACGAACATCTCATGAGCCAGACGTCCTCCAGCGATAGCAACG

                *         380         *         400
3b  :  GCCACGAGGACTCTCGCCGGTGGAGATCCGAAATTTGTAAAAATG  :   405
3a  :  GCCACGAGGACTCTCGCCGGTGGAGATCCGAAATTTGTAAAAATG  :   405
       GCCACGAGGACTCTCGCCGGTGGAGATCCGAAATTTGTAAAAATG

            *         420         *         440         *
3b  :  GTAACAAGACCAAAACCAAAAGTCATTTTCCTCATG---------  :   441
3a  :  GTAACAAGACCAAAACCAAAAGTCATTTTCCTCATGCATCTGAAG  :   450
       GTAACAAGACCAAAACCAAAAGTCATTTTCCTCATGCATCTGAAG
```

```
            460         *       480         *
3b : ------------------------------------------- :   -
3a : AACAAAAGGAAACATCTGTCACAGAAATGCAGAGTAGTCCCCCAG :  495
     AACAAAAGGAAACATCTGTCACAGAAATGCAGAGTAGTCCCCCAG

       500         *       520         *       540
3b : ------------------------------------------- :   -
3a : CTCAGGCGAAAGCTGTTCCAGCCACAGAAGAGGACAGCCTGGGGG :  540
     CTCAGGCGAAAGCTGTTCCAGCCACAGAAGAGGACAGCCTGGGGG

             *       560         *       580
3b : ------------------------------------------- :   -
3a : CCAGTGTGCCCAGGACTGGCGTCCCAGAGGAGCTGGCCTGCAAGG :  585
     CCAGTGTGCCCAGGACTGGCGTCCCAGAGGAGCTGGCCTGCAAGG

         *       600         *       620         *
3b : ------------------------------------------- :   -
3a : ACCAGCCCGCCTGCTCCTACCAGCAGATCAGCTGCCTGGATGGCG :  630
     ACCAGCCCGCCTGCTCCTACCAGCAGATCAGCTGCCTGGATGGCG

            640         *       660         *
3b : ------------------------------------------- :   -
3a : TCCTCAGGTACCTGGAGAGCTGCAGCGAGGCCACCACCCTGAAGG :  675
     TCCTCAGGTACCTGGAGAGCTGCAGCGAGGCCACCACCCTGAAGG

       680         *       700         *       720
3b : ------------------------------------------- :   -
3a : GGAAGTGCGAGTTCCCAGCACACCTCCCGTCACCAAAGGCCACCA :  720
     GGAAGTGCGAGTTCCCAGCACACCTCCCGTCACCAAAGGCCACCA

             *       740         *       760
3b : ------------------------------------------- :   -
3a : CCAGCCCTGGGCTGCACACCGCAGGGGCAGCGTCGCCCTCCAAGG :  765
     CCAGCCCTGGGCTGCACACCGCAGGGGCAGCGTCGCCCTCCAAGG

         *       780         *       800         *
3b : ------------------------------------------- :   -
3a : TGAGCAGCCGCAGGGAAGTCAGTGCCCGCCTGAGCTCGCTGGCAC :  810
     TGAGCAGCCGCAGGGAAGTCAGTGCCCGCCTGAGCTCGCTGGCAC

            820         *       840         *
3b : ------------------------------------------- :   -
3a : TGCCTGGCAAGGCTGAGAGCGTAGTGTCCCTCACCAGCCAGTGCA :  855
     TGCCTGGCAAGGCTGAGAGCGTAGTGTCCCTCACCAGCCAGTGCA

       860         *       880         *       900
3b : ------------------------------------------- :   -
3a : GCTATAGCAGCACCATTGTCCACGTGGGGGACAAAAAGCCACAGC :  900
     GCTATAGCAGCACCATTGTCCACGTGGGGGACAAAAAGCCACAGC
```

```
            *       920         *       940
3b : ---------------------------------------- :   -
3a : CTGAGTCAGAGATGGCGGAGGATGCGGCCAGTGGGCCCGAGGCCC : 945
     CTGAGTCAGAGATGGCGGAGGATGCGGCCAGTGGGCCCGAGGCCC

         *       960         *       980         *
3b : ---------------------------------------- :   -
3a : TGGACGGCCTGCCCCACAGCCTCAGCCAGGAGAAGGGGCCCCTGC : 990
     TGGACGGCCTGCCCCACAGCCTCAGCCAGGAGAAGGGGCCCCTGC

           1000        *      1020        *
3b : ---------------------------------------- :   -
3a : AGAAGCTGGGCCTCACCAAGGAGGTGCTGGCCGCGCACACGCAGA : 1035
     AGAAGCTGGGCCTCACCAAGGAGGTGCTGGCCGCGCACACGCAGA

      1040        *      1060        *      1080
3b : ---------------------------------------- :   -
3a : GGGAGGAGCAGAGCTTCCTGCAGAAGTTCCGCGGAGCCAGGAGAC : 1080
     GGGAGGAGCAGAGCTTCCTGCAGAAGTTCCGCGGAGCCAGGAGAC

            *      1100        *      1120
3b : ---------------------------------------- :   -
3a : TCGGGATTCTCCAGGCTCACTGCCATTACTACTCACGGGACAGAT : 1125
     TCGGGATTCTCCAGGCTCACTGCCATTACTACTCACGGGACAGAT

         *      1140        *      1160        *
3b : ---------------------------------------- :   -
3a : CTGGGGGGCCCCCGAGCGAGCGCACTGCCCCTGGACTGAGAAGTG : 1170
     CTGGGGGGCCCCCGAGCGAGCGCACTGCCCCTGGACTGAGAAGTG

           1180        *      1200        *
3b : ---------------------------------------- :   -
3a : CTTCTGGGATAGACTCGTCCTGGAGGAAACCAGGGAAGGGCAGAA : 1215
     CTTCTGGGATAGACTCGTCCTGGAGGAAACCAGGGAAGGGCAGAA

      1220        *      1240        *      1260
3b : ---------------------------------------- :   -
3a : AGCTCAAGTCCAGAAGGGCCAAGCCTCGGGGGACCCCCCAGAGCA : 1260
     AGCTCAAGTCCAGAAGGGCCAAGCCTCGGGGGACCCCCCAGAGCA

            *      1280        *      1300
3b : ---------------------------------------- :   -
3a : CAGGGCCTGGGGCGCCGGAGCCCCCACGGCGCCCACTCATCAGCC : 1305
     CAGGGCCTGGGGCGCCGGAGCCCCCACGGCGCCCACTCATCAGCC

         *      1320        *      1340        *
3b : ---------------------------------------- :   -
3a : TGAGTGCCATGGCCTGGTCGCCCTCGGACACGTCCCAGTCTAGCT : 1350
     TGAGTGCCATGGCCTGGTCGCCCTCGGACACGTCCCAGTCTAGCT
```

```
          1360         *      1380         *
3b : ---------------------------------------- :    -
3a : GCCCTGCCGTGGCCTTCCCCACCTCAGTGCCCACGTGCCCGCTGC : 1395
     GCCCTGCCGTGGCCTTCCCCACCTCAGTGCCCACGTGCCCGCTGC

      1400         *      1420         *      1440
3b : ---------------------------------------- :    -
3a : CCGTGTTCCCGGCGCCAGGCCCCTTGGTGTCGCCGCCTGCAGATC : 1440
     CCGTGTTCCCGGCGCCAGGCCCCTTGGTGTCGCCGCCTGCAGATC

               *      1460         *      1480
3b : ---------------------------------------- :    -
3a : CCCACGCCAGCCTTGCCATGCCCACTCTGCCTGTGGACACCCAGC : 1485
     CCCACGCCAGCCTTGCCATGCCCACTCTGCCTGTGGACACCCAGC

          *      1500         *      1520         *
3b : ---------------------------------------- :    -
3a : CCGAGTTCGCAGTCCAGCCCCTGCCGTTTGCCGCCCCCTTGGCCC : 1530
     CCGAGTTCGCAGTCCAGCCCCTGCCGTTTGCCGCCCCCTTGGCCC

          1540         *      1560         *
3b : --------GCCCTCGTGCTCCCCACCTGTCCCTTCCCACCGGTGG :  478
3a : CGGTCATGGCCCTCGTGCTCCCCACCTGTCCCTTCCCACCGGTGG : 1575
     CGGTCATGGCCCTCGTGCTCCCCACCTGTCCCTTCCCACCGGTGG

      1580         *      1600         *      1620
3b : CCCCAGGCCTCCCCCAGGCCTTCTTCCCTGGCCAGCCTCACTTTC :  523
3a : CCCCAGGCCTCCCCCAGGCCTTCTTCCCTGGCCAGCCTCACTTTC : 1620
     CCCCAGGCCTCCCCCAGGCCTTCTTCCCTGGCCAGCCTCACTTTC

               *      1640         *      1660
3b : CAGGCCACCCCACGCTCCCCTCCGAGATGCGTCCTGCCCCACAGC :  568
3a : CAGGCCACCCCACGCTCCCCTCCGAGATGCGTCCTGCCCCACAGC : 1665
     CAGGCCACCCCACGCTCCCCTCCGAGATGCGTCCTGCCCCACAGC

          *      1680         *      1700         *
3b : CTGAGCTCCCCCATCAGACCTCACTCCCAGGACAGCCCTGCGCTT :  613
3a : CTGAGCTCCCCCATCAGACCTCACTCCCAGGACAGCCCTGCGCTT : 1710
     CTGAGCTCCCCCATCAGACCTCACTCCCAGGACAGCCCTGCGCTT

          1720         *      1740         *
3b : GCCCAGCCACCCCACCGTCAGCCATGGTGGCCACGGGCAGGGCCT :  658
3a : GCCCAGCCACCCCACCGTCAGCCATGGTGGCCACGGGCAGGGCCT : 1755
     GCCCAGCCACCCCACCGTCAGCCATGGTGGCCACGGGCAGGGCCT

      1760         *      1780         *      1800
3b : CCCCGCCGCTCTTCCAGTCCCGAGGCAGCTCGCCCCTGCAGCTCA :  703
3a : CCCCGCCGCTCTTCCAGTCCCGAGGCAGCTCGCCCCTGCAGCTCA : 1800
     CCCCGCCGCTCTTCCAGTCCCGAGGCAGCTCGCCCCTGCAGCTCA
```

```
                *      1820          *      1840
3b : ACCTGCTGCAGATGGAGGAGCCCCCCGAGGGCAGCGCCGGGGCTG :  748
3a : ACCTGCTGCAGATGGAGGAGCCCCCCGAGGGCAGCGCCGGGGCTG : 1845
     ACCTGCTGCAGATGGAGGAGCCCCCCGAGGGCAGCGCCGGGGCTG

            *      1860          *      1880          *
3b : CGGCACCCCTGGGCACTCCTGGAGCAGCAGCCGCCAGGCCGGACT :  793
3a : CGGCACCCCTGGGCACTCCTGGAGCAGCAGCCGCCAGGCCGGACT : 1890
     CGGCACCCCTGGGCACTCCTGGAGCAGCAGCCGCCAGGCCGGACT

          1900          *      1920          *
3b : GCACCCCTGGCGCTTCTCGGGACCCGCAGCCGAAGGCACCCCCAA :  838
3a : GCACCCCTGGCGCTTCTCGGGACCCGCAGCCGAAGGCACCCCCAA : 1935
     GCACCCCTGGCGCTTCTCGGGACCCGCAGCCGAAGGCACCCCCAA

      1940          *      1960          *      1980
3b : CGCGAGACGAGCCCCTAGATGCACAGGAGAGTGACGCCCTGTCTG :  883
3a : CGCGAGACGAGCCCCTAGATGCACAGGAGAGTGACGCCCTGTCTG : 1980
     CGCGAGACGAGCCCCTAGATGCACAGGAGAGTGACGCCCTGTCTG

                *      2000          *      2020
3b : TGTCCAGCGACCTGCTCAACCTCCTGCTGAGTGAGGACGTCTGCT :  928
3a : TGTCCAGCGACCTGCTCAACCTCCTGCTGAGTGAGGACGTCTGCT : 2025
     TGTCCAGCGACCTGCTCAACCTCCTGCTGAGTGAGGACGTCTGCT

            *      2040          *      2060          *
3b : CTGCCGCGGGGTCAGCCCTCTCTGGGAGTGGGGCCTCTGCCGCCT :  973
3a : CTGCCGCGGGGTCAGCCCTCTCTGGGAGTGGGGCCTCTGCCGCCT : 2070
     CTGCCGCGGGGTCAGCCCTCTCTGGGAGTGGGGCCTCTGCCGCCT

          2080          *      2100          *
3b : CGGATTCTCTGGGCTGCGGTGCATCCCAAAGTGGGGCAGGCAGCA : 1018
3a : CGGATTCTCTGGGCTGCGGTGCATCCCAAAGTGGGGCAGGCAGCA : 2115
     CGGATTCTCTGGGCTGCGGTGCATCCCAAAGTGGGGCAGGCAGCA

      2120          *      2140          *      2160
3b : GTGACACCAGTCGTACCAGCAAATACTTTGGAAGCATCGACTCTT : 1063
3a : GTGACACCAGTCGTACCAGCAAATACTTTGGAAGCATCGACTCTT : 2160
     GTGACACCAGTCGTACCAGCAAATACTTTGGAAGCATCGACTCTT

                *      2180          *      2200
3b : CAGAGAATAATCACAGAGCAGAAATGAACAGGGACAAGGAGCAGA : 1108
3a : CAGAGAATAATCACAGAGCAGAAATGAACAGGGACAAGGAGCAGA : 2205
     CAGAGAATAATCACAGAGCAGAAATGAACAGGGACAAGGAGCAGA

            *      2220          *      2240          *
3b : GTGAACACTTCATCAAGTACGTCCTTCAGGACCCCATCTGGCTGC : 1153
3a : GTGAACACTTCATCAAGTACGTCCTTCAGGACCCCATCTGGCTGC : 2250
     GTGAACACTTCATCAAGTACGTCCTTCAGGACCCCATCTGGCTGC
```

```
          2260         *      2280         *
3b : TGATGGCGGACGCGGACGACAGCGTCATGATGACCTACCAGCTGC : 1198
3a : TGATGGCGGACGCGGACGACAGCGTCATGATGACCTACCAGCTGC : 2295
     TGATGGCGGACGCGGACGACAGCGTCATGATGACCTACCAGCTGC

      2300         *      2320         *      2340
3b : CTTCCCGGGATCTGGAGTCAGTCTTGAAGGAGGACAGGGAGAAGC : 1243
3a : CTTCCCGGGATCTGGAGTCAGTCTTGAAGGAGGACAGGGAGAAGC : 2340
     CTTCCCGGGATCTGGAGTCAGTCTTGAAGGAGGACAGGGAGAAGC

             *      2360         *      2380
3b : TGAAGCTACTGCAGAGGTCCCAGCCCTGGTTCACTGAGGGCCAGA : 1288
3a : TGAAGCTACTGCAGAGGTCCCAGCCCTGGTTCACTGAGGGCCAGA : 2385
     TGAAGCTACTGCAGAGGTCCCAGCCCTGGTTCACTGAGGGCCAGA

         *      2400         *      2420         *
3b : GGCAGGAGCTGCGTGAGGTCCACCCATGGGTCCAGACAGGTGGCC : 1333
3a : GGCAGGAGCTGCGTGAGGTCCACCCATGGGTCCAGACAGGTGGCC : 2430
     GGCAGGAGCTGCGTGAGGTCCACCCATGGGTCCAGACAGGTGGCC

          2440         *      2460         *
3b : TGCCCGCGGCCATTGATGTGGCGGAATGTGTTTACTGTGACAGCG : 1378
3a : TGCCCGCGGCCATTGATGTGGCGGAATGTGTTTACTGTGACAGCG : 2475
     TGCCCGCGGCCATTGATGTGGCGGAATGTGTTTACTGTGACAGCG

      2480         *      2500         *      2520
3b : AAGAGAAAGGCAGTGTTTGCGTGCCATTTGAGGGAGACATTCCTT : 1423
3a : AAGAGAAAGGCAGTGTTTGCGTGCCATTTGAGGGAGACATTCCTT : 2520
     AAGAGAAAGGCAGTGTTTGCGTGCCATTTGAGGGAGACATTCCTT

             *      2540         *      2560
3b : CCCTGGGACTCAGCGACATGTCCGACACCAAAGAAGAGGAAAGCG : 1468
3a : CCCTGGGACTCAGCGACATGTCCGACACCAAAGAAGAGGAAAGCG : 2565
     CCCTGGGACTCAGCGACATGTCCGACACCAAAGAAGAGGAAAGCG

         *      2580         *      2600
3b : GACACCCCCCGAGGCACAGGGAGAAGGAGCGGACGTAA : 1506
3a : GACACCCCCCGAGGCACAGGGAGAAGGAGCGGACGTAA : 2603
     GACACCCCCCGAGGCACAGGGAGAAGGAGCGGACGTAA
```

D4. Table showing the prediction of region showing intronic retention and multiple exon deletion

F.pal Per2v predicting the region of possible intronic retention and multiple exon deletion

INTRONIC RETENTION BETWEEN 10th and 11th EXON

ORGANISM	POSITION OF 9th EXON	POSITION OF 10th EXON	POSITION OF THE LAST AND FIRST BASE OF EXON FLANKING THE 503bp INSERTION **ACAARAAGA- intron-TCCTRCAGKC**
Homo sapiens	25509-25615	26223-26376	25615 and 26223
Rattus rattus	22964-23070	23548-23701	23070 and 23548
Mus musculus	23573-23679	24159-24312	23679 and 24159

MULTIPLE EXON DELETION: FROM 13th TO 18th

ORGANISM	EXON ARCHITECTURE OF mRNA-CDS CONSTITUTING EXON 13 to EXON 18	POSITION OF THE DELETION **CAKCTGAAGRAC….GCBCCKSTCRTG**
Homo sapiens	28514-28598 29923-30070 30167-30291 31481-31645 32656-32910 34865-35655	28566 …..to……35257
Rattus rattus	25927-26011 27936-28098 28189-28295 29757- 29921 31425-31664 31774-35570	25979……to……35172
Mus musculus	26437-26521 28189-28351 28449-28555 29781-29945 31351-31590 34843-35639	26489……to……35241

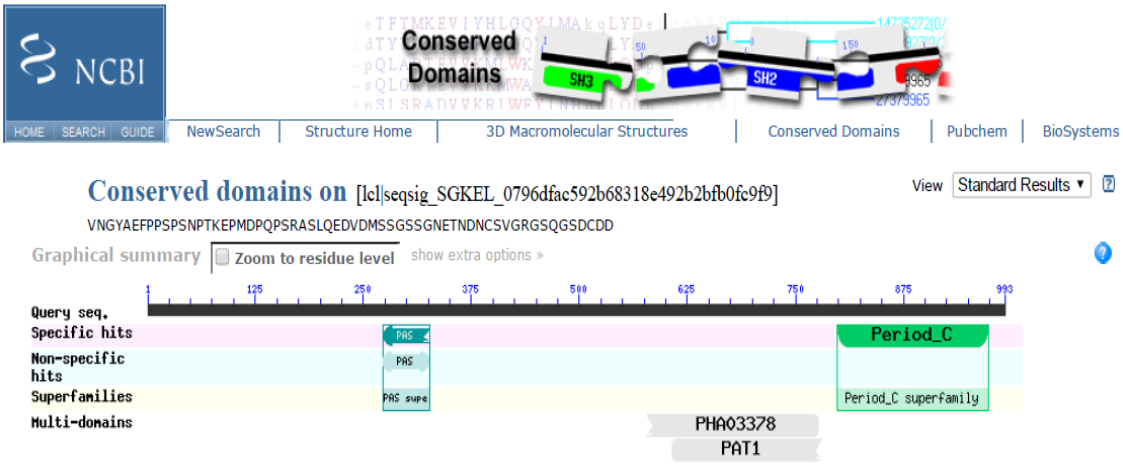
Translate Tool - Results of translation

Open reading frames are highlighted in red. The PTC is highlighted

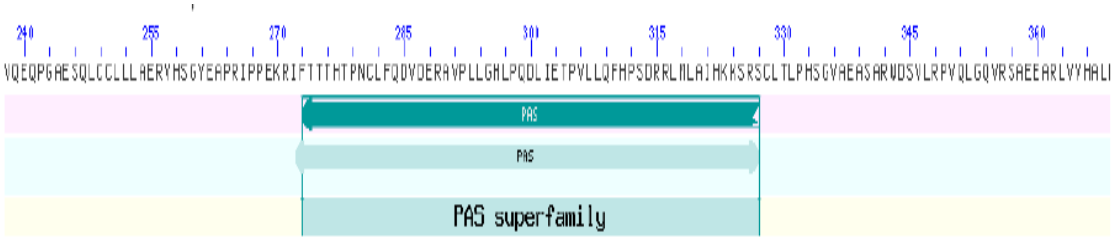
5'3'Frame1

V N G Y A E F P P S P S N P T K E P **Met** D P Q P S R A S L Q E D V D **Met** S S G
S S G N E T N D N C S V G R G S Q G S D C D D S G K E L G **Met** L V E P A D T
R E S P G A F H L **Met** **Met** T E P E R N P S T S G C S S E Q S A K A D T H R E L
I R T L R E L K V H L P A D K Q A K G K A S T L A T L K Y A L R S V K Q V K
A N E E Y F Q L L **Met** S S D S P P W G A G V P S Y T V E Q V E G V T S E Y L L
K N A D **Met** F A V A V S L V T G R I L Y I S N Q V T S I F H C K R D A F G D A
K F V E F Q G P Q D V S V F H S F T T P C K L P P W S A C G G V D S F T Q E
C **Met** E E K S F F C R V G V G Q N H E N G I R Y Q P F R **Met** T A Y L A Q V Q
E Q P G A E S Q L C C L L L A E R V H S G Y E A P R I P P E K R I F T T T H T P
N C L F Q D V D E R A V P L L G H L P Q D L I E T P V L L Q F H P S D R R
L **Met** L A I H K K S R S C L T L P H S G V A E A S A R W D S V L R P V Q L G
Q V R S A E E A R L V V H A L L P A R A R H Q P G R D Q E G I C G V R C P L P
V A **Stop** V S D V S F L L L A S **Met** A G G C G G D H F R G V Y D **Met** C P H H
A Q P S Q T G D T L S R D V A T L W E A T P G C N S E Q L Q W L E D L S A P
R Q C A C D P R L L H P W C L A S A L L T I C S S L P **Stop** S C S P V G S L S T I
L P C A S A R G T G S T S L W T P A G P A S S T R G A G R S P S S S G G T E S
G W A L **Stop** T R T C S Q H P R A Q S R R P F S P A F R S S R S R S I G C C C S
R S P T A A P V A T G A W A A T G P T N I S **Stop** A R R P P A I A T A T R T L A
G G D P K F V K **Met** V T R P K P K V I F L **Met** A L V L P T C P F P P V A P G L
P Q A F F P G Q P H F P G H P T L P S E **Met** R P A P Q P E L P H Q T S L P G Q P
C A C P A T P P S A **Met** V A T G R A S P P L F Q S R G S S P L Q L N L L
Q **Met** E E P P E G S A G A A A P L G T P G A A A A R P D C T P G A S R D P Q P
K A P P T R D E P L D A Q E S D A L S V S S D L L N L L L S E D V C S A A G S
A L S G S G A S A A S D S L G C G A S Q S G A G S S D T S R T S K Y F G S I D
S S E N N H R A E **Met** N R D K E Q S E H F I K Y V L Q D P I W L L **Met** A D A
D D S V **Met** **Met** T Y Q L P S R D L E S V L K E D R E K L K L L Q R S Q P W F
T E G Q R Q E L R E V H P W V Q T G G L P A A I D V A E C V Y C D S E E K G
S V C V P F E G D I P S L G L S D **Met** S D T K E E E S G H P P R H R E K E R
T **Stop**

Conserved Domain Search-NCBI - *F.pal Per2v* retains the PAS domains



Expanded view of PAS region



2. (a) Cloning and Localization of SCN expressed Vasoactive *Intestinal peptide*

(b) Localization of *Per2* transcripts in SCN using *Vip* as a marker for core

2(a) Cloning and Localization of SCN expressed Vasoactive *Intestinal peptide (Vip)*

PCR amplification and Cloning of *Vip*

The pre-pro *Vip* encodes both VIP and a peptide-histidine Isoleucine (PHI) residue which are co-expressed in most of the tissues. Hence degenerate primers were designed to amplify the entire pre-pro polypeptide from the *F. palmarum* SCN (Fig. 28).

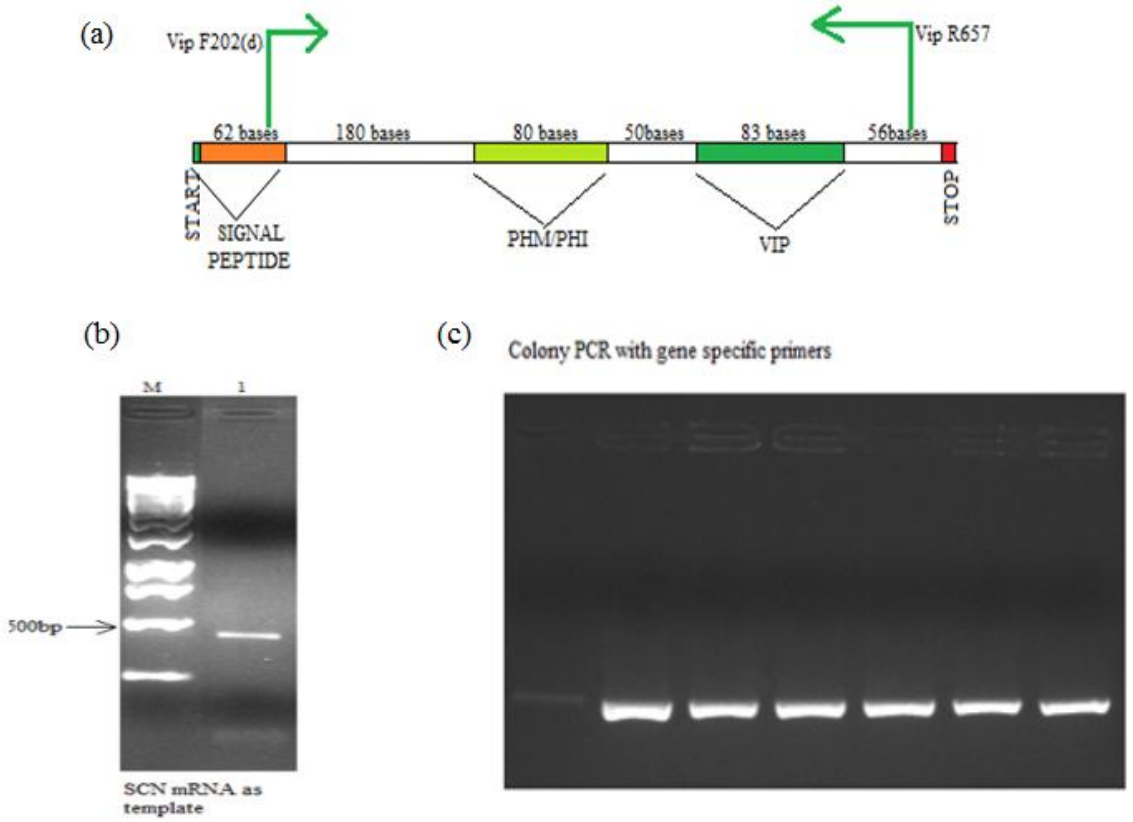
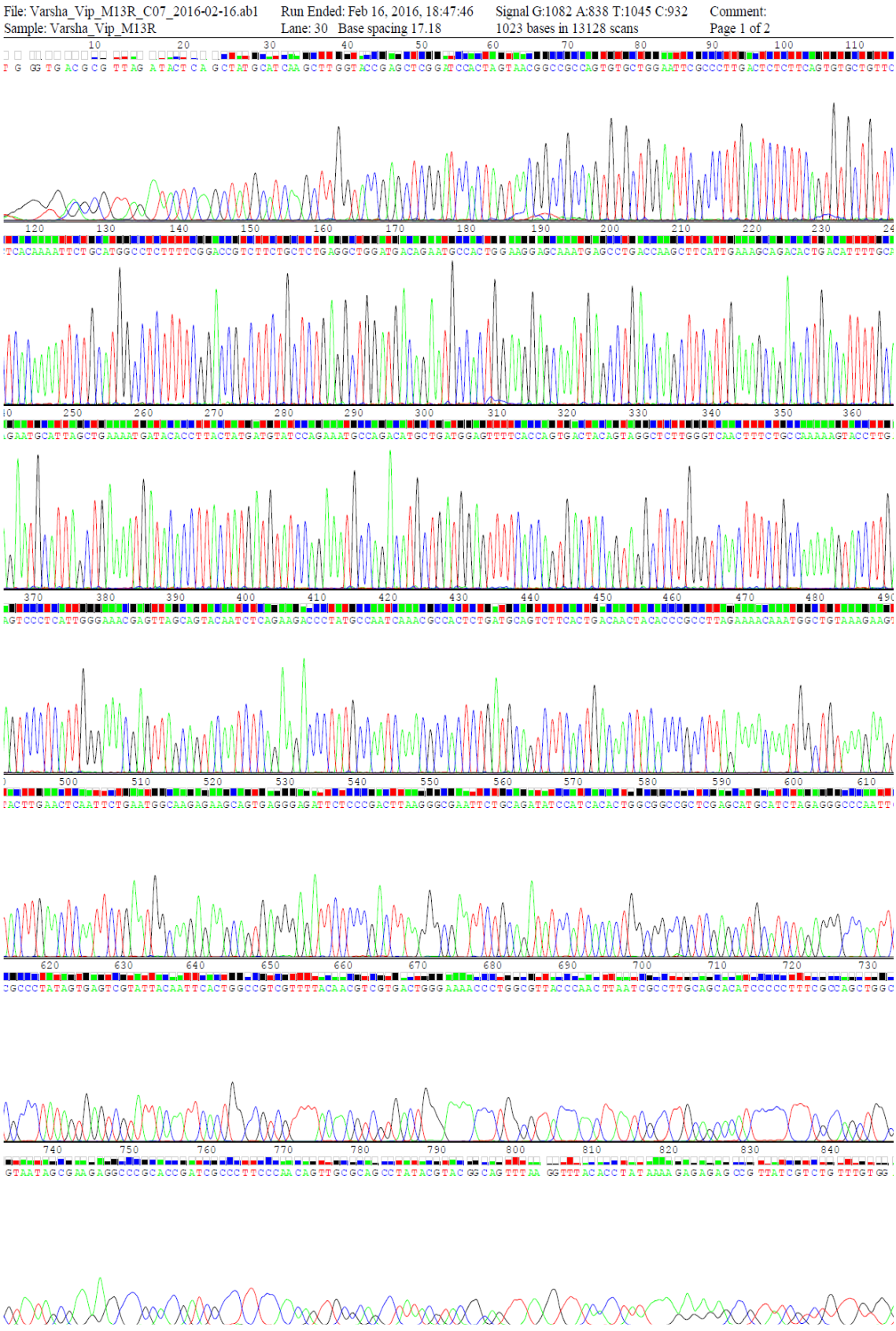


Fig.28. Sense and antisense primer position depicted on pre-pro *Vip* polypeptide. (b) Agarose gel with the PCR product just below the 500bp range. (c) Colony PCR results from 6 different positive colonies after TA cloning.

Sequence analysis identified the gene as pre-pro *Vip* polyeptide, the *F.palmarum Vip* showed more than 80% similarity to previously reported nucleotide sequences in mammals (C27). The PHI domain was 96% similar to primate and alpine lemur (Scuridae) sequences than to mouse or rat. A similar pattern of similarity was observed in the *Vip* domain as well. The Spacer region was more variant and showed 80 to 85% resemblance to *Mus musculus* and *Homo sapiens* respectively, whereas the region showed 92% similarity to alpine marmot. Overall the pre-pro polypeptide showed significant similarity to primates and alpine marmot over rat or mouse.

C27. *F.palmarum Vip* in TA plasmid. Primer M13F



>Varsha_Vip_M13R sequence exported from Varsha_Vip_M13R_C07_2016-0216.ab1

Green- Forward and Reverse *Vip* primer sequence, Yellow- PHI region, Red- *Vip* region

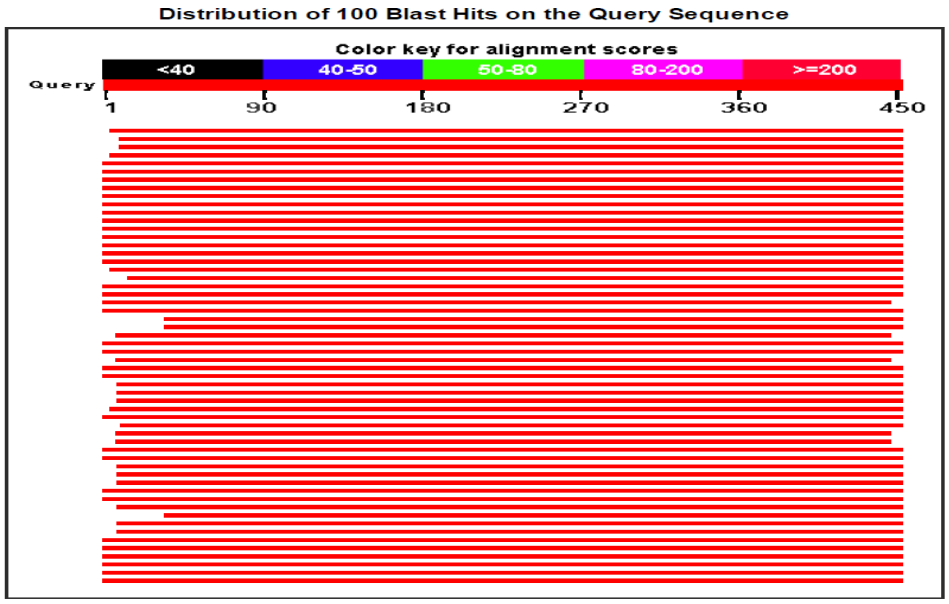
TGGGTGACGCGTTAGATACTCAGCTATGCATCAAGCTTGGTACCGAGCTCG
GATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGACTCTCTTC
AGTGTGCTGTTCTCACAAAATTCTGCATGGCCTCTTTTCGGACCGTCTTCTGC
TCTGAGGCTGGATGACAGAATGCCACTGGAAGGAGCAAATGAGCCTGACCA
AGCTTCATTGAAAGCAGACACTGACATTTTGCAGAATGCATTAGCTGAAAA
TGATACACCTTACTATGATGTATCCAGAAATGCCAGACATGCTGATGGAGTT
TTCACCAGTGACTACAGTAGGCTCTTGGGTCAACTTTCTGCCAAAAAGTACC
TTGAGTCCCTCATTGGGAAACGAGTTAGCAGTACAATCTCAGAAGACCCTA
TGCCAATCAAACGCCACTCTGATGCAGTCTTCACTGACAACTACACCCGCCT
TAGAAAACAAATGGCTGTAAAGAAGTACTTGAACTCAATTCTGAATGGCAA
GAGAAGCAGTGAGGGAGATTCTCCCGACTT

NCBI BLAST report for the insert. First 20 his are enlisted below.

Nucleotide Sequence (453 letters)

RID	PVZFTRDC014 (Expires on 06-26 11:43 am)		
Query ID	lcl|Query_234677	**Database Name**	nr
Description	None	**Description**	Nucleotide collection (nt)
Molecule type	nucleic acid	**Program**	BLASTN 2.4.0+
Query Length	453		

Graphic Summary



Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: Marmota marmota marmota vasoactive intestinal peptide (Vip), transcript variant X2, mRNA	658	658	99%	0.0	93%	XM_015498803.1

PREDICTED: Ictidomys tridecemlineatus vasoactive intestinal peptide (Vip), transcript variant X1, mRNA	654	654	98%	0.0	93%	XM_005321382.2
PREDICTED: Ictidomys tridecemlineatus vasoactive intestinal peptide (Vip), transcript variant X2, mRNA	634	634	98%	5e-178	93%	XM_013356895.1
PREDICTED: Marmota marmota marmota vasoactive intestinal peptide (Vip), transcript variant X1, mRNA	580	580	99%	6e-162	90%	XM_015498799.1
PREDICTED: Microcebus murinus vasoactive intestinal peptide (VIP), transcript variant X1, mRNA	571	571	100%	4e-159	89%	XM_012748154.1
PREDICTED: Microcebus murinus vasoactive intestinal peptide (VIP), transcript variant X2, mRNA	556	556	100%	1e-154	89%	XM_012748155.1
Homo sapiens vasoactive intestinal peptide (VIP), transcript variant 1, mRNA	544	544	100%	9e-151	88%	NM_003381.3
Human vasoactive intestinal peptide (VIP) mRNA, complete cds	544	544	100%	9e-151	88%	M36634.1
Homo sapiens cDNA, FLJ94595	538	538	100%	4e-149	88%	AK313950.1
PREDICTED: Homo sapiens vasoactive intestinal peptide (VIP), transcript variant X1, mRNA	529	529	100%	2e-146	88%	XM_006715562.3
Synthetic construct Homo sapiens clone ccsbBroadEn_01772 VIP gene, encodes complete protein	529	529	100%	2e-146	88%	KJ892378.1
Homo sapiens vasoactive intestinal peptide (VIP), transcript variant 2, mRNA	529	529	100%	2e-146	88%	NM_194435.2
Synthetic construct Homo sapiens clone IMAGE:100071241; CCSB006418_02 vasoactive intestinal peptide (VIP) gene, encodes complete protein	529	529	100%	2e-146	88%	HQ447888.1
Synthetic construct Homo sapiens gateway clone IMAGE:100018534 3' read VIP mRNA	529	529	100%	2e-146	88%	CU679496.1
Synthetic construct Homo sapiens gateway clone IMAGE:100018534 5' read VIP mRNA	529	529	100%	2e-146	88%	CU679495.1
Homo sapiens vasoactive intestinal peptide, mRNA (cDNA clone MGC:13587 IMAGE:4294373), complete cds	529	529	100%	2e-146	88%	BC009794.1
PREDICTED: Pan troglodytes vasoactive intestinal peptide (VIP), transcript variant X1, mRNA	527	527	100%	9e-146	88%	XM_527541.4
PREDICTED: Peromyscus maniculatus bairdii vasoactive intestinal peptide (Vip), transcript variant X1, mRNA	525	525	99%	3e-145	88%	XM_006978083.2
PREDICTED: Ceratotherium simum simum vasoactive intestinal peptide (LOC101405118), mRNA	523	523	96%	1e-144	88%	XM_004440686.2

2(b) Localization of *Per2* transcripts in SCN using *Vip* as a marker for core

Generation of *in situ* hybridization probes by *in vitro* transcription.

Gene specific primers with T7/T3 polymerase promoters were used to generate PCR products from TA plasmids carrying the *F. palmarum, Per2, Per2v* and *Vip* cDNA inserts.

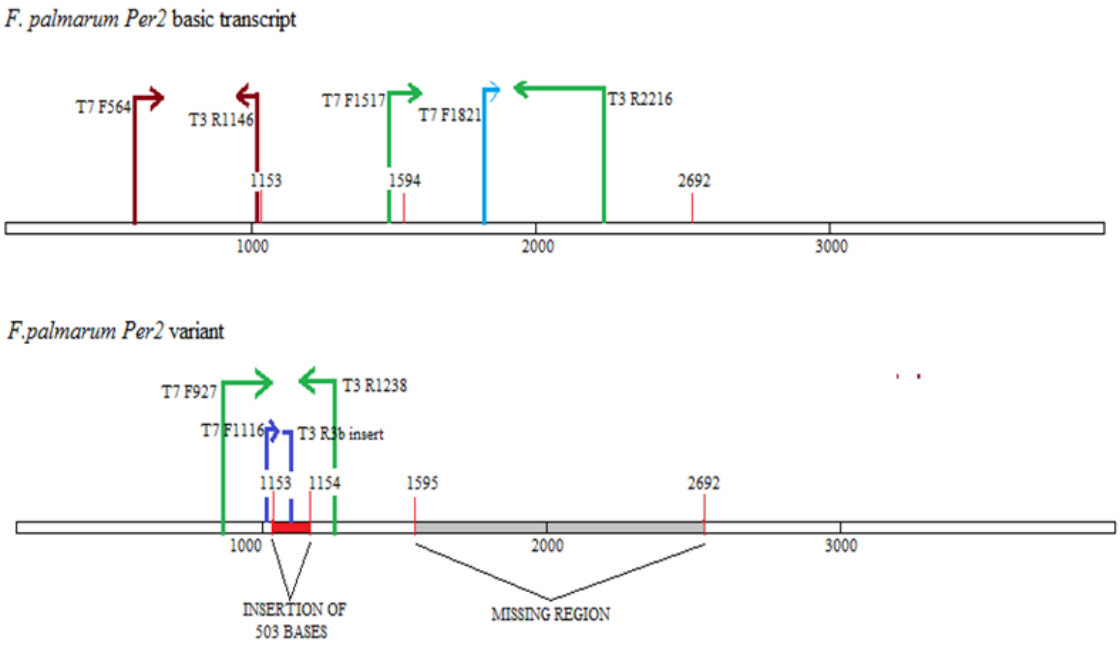


Fig.29. Diagram showing the primers that are designed to distinguish between the variants of *Per2*

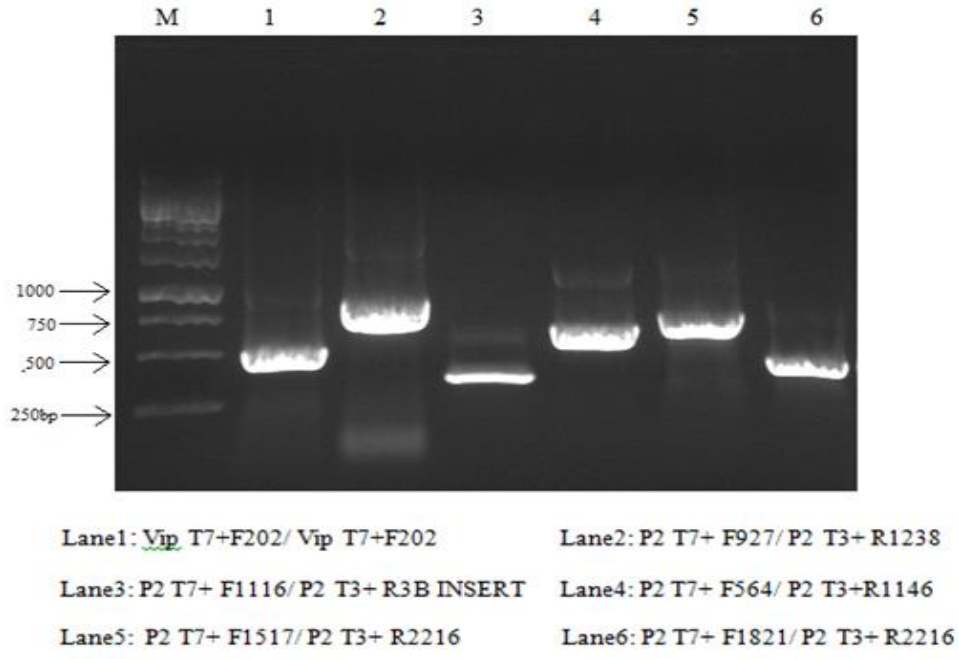
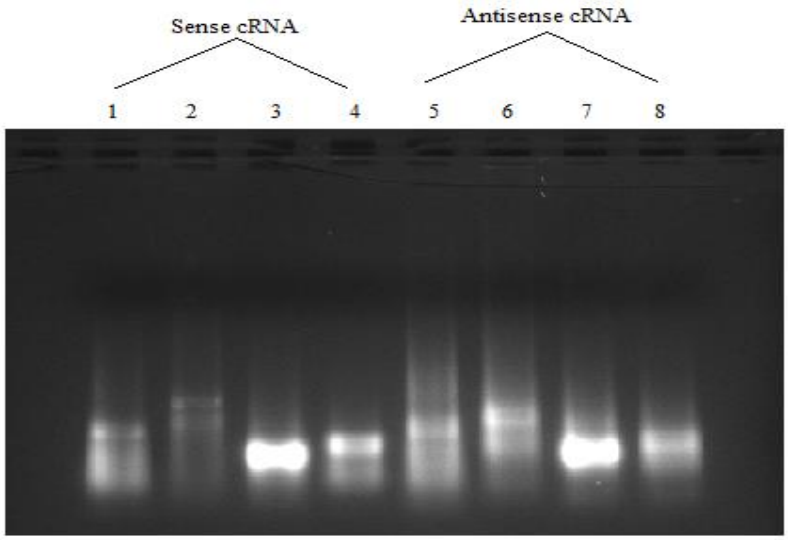


Fig.30. Agarose gel with specific amplicons that will be substrates for in vitro transcription

The amplicons were excised and extracted using kit method (Qiagen), an aliquot of the same was send for sequencing before using it for generating digoxigenin labelled cRNA by in-vitro transcription. In vitro transcription using T7 polymerase generated sense probes whereas T3 polymerase generated antisense probes. The probe integrity was checked by gel electrophoresis before using on sections (Fig. 31).



Lanes 1 and 5- Sense and Antisense for *F.pal Vip*
Lanes 2 and 6- Sense and Antisense for *F.pal Per2*(T7F1517/T3R2216 region)
Lanes 3 and 7- Sense and Antisense for *F.pal Per2v*(T7 F1116/T3 R3B insert region)
Lanes 4 and 8- Sense and Antisense for *F.pal Per2v*(T7F1041 E8-E9/R3B insert region)

* The PCR image of T7F1041 E8-E9/R3B insert is not depicted in Fig.30
* The probes in lanes 3 and 7 were not used for the present work.

Fig.31. Integrity of cRNA probes confirmed by gel electrophoresis

In situ hybridization was performed on free floating 40µm cryosections, taken across the entire extent of SCN. Alternate sections were probed and examined for, *Vip, Per2,* and *Per2v* immunoreactivity. Analysis of the images is in progress.

3. Effect of various photoperiods on Clock gene expression (*Per1, Per2, Per3, Cry1, Cry2 and Bmal1*)

Individual primer pairs for each clock gene were checked for their specificity by conventional PCR. The amplicons were excised from the agarose gel, DNA was extracted and each sample was sent for sequencing to confirm identity of the gene. Sequence analysis and BLAST report for individual gene is attached as separate document (C28 to C31).

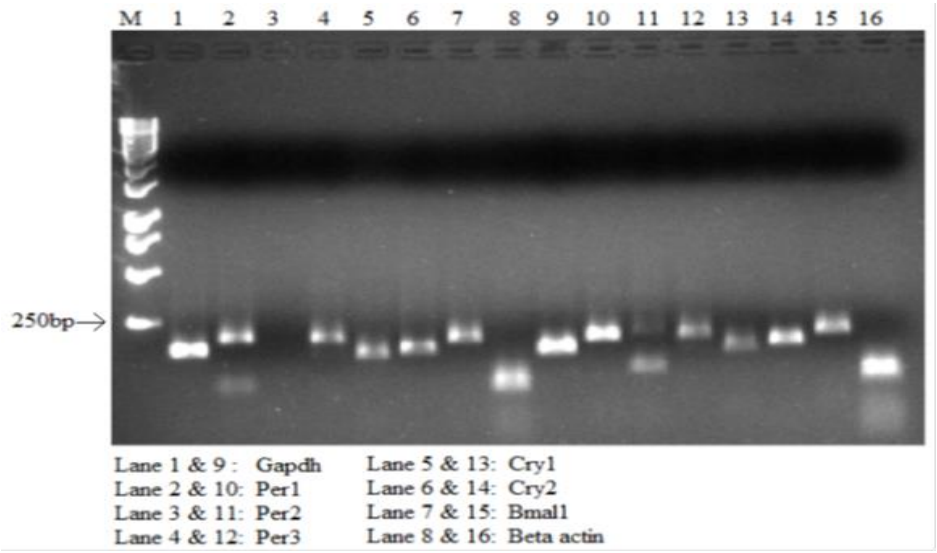
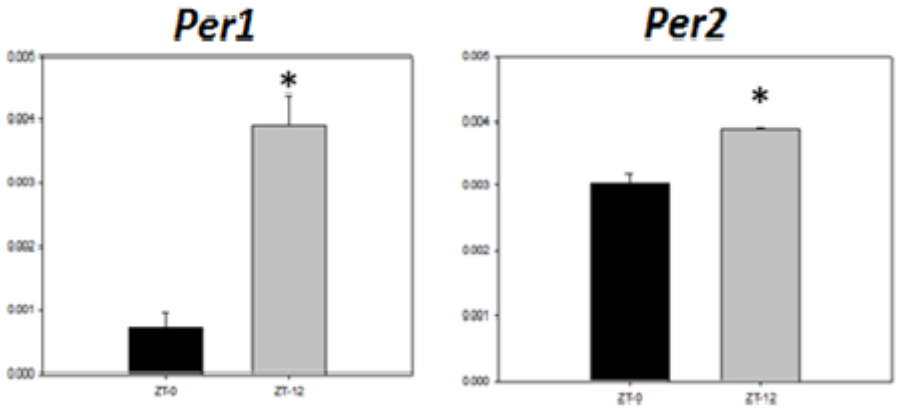


Fig.32 PCR amplification of *F. palmarum* clock genes, agarose gel showing single specific amplification for each gene.

Relative quantification of clock gene expression was done using *Gapdh* as endogenous control



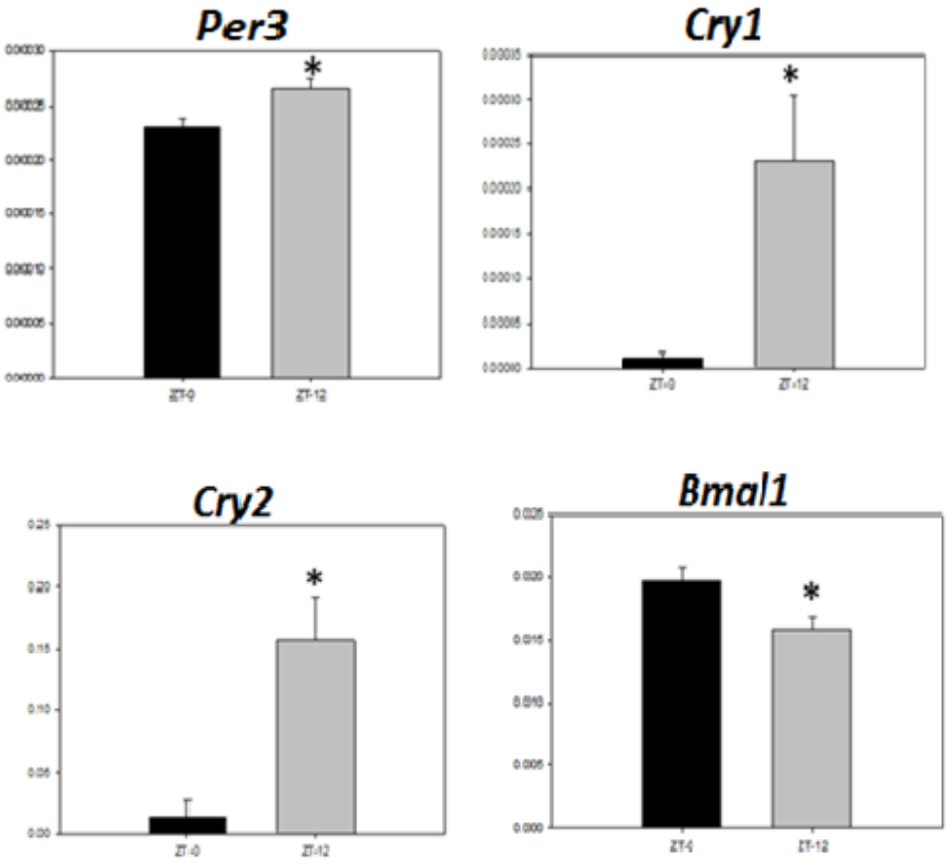
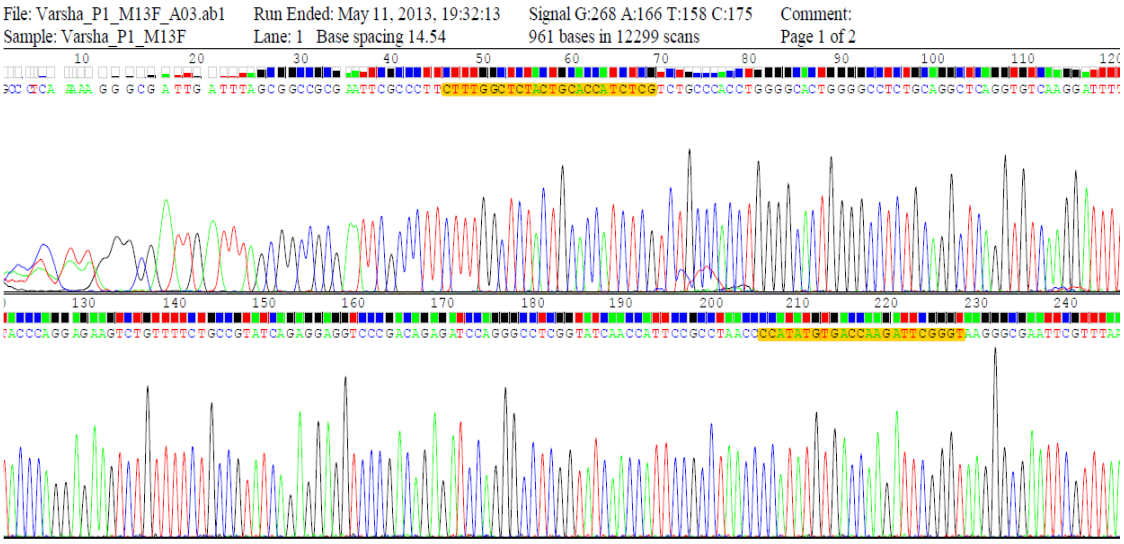


Fig.33. Histogram showing the expression of clock genes at ZT0 and ZT 12

X axis- Relative gene expression. Y axis-ZT time points

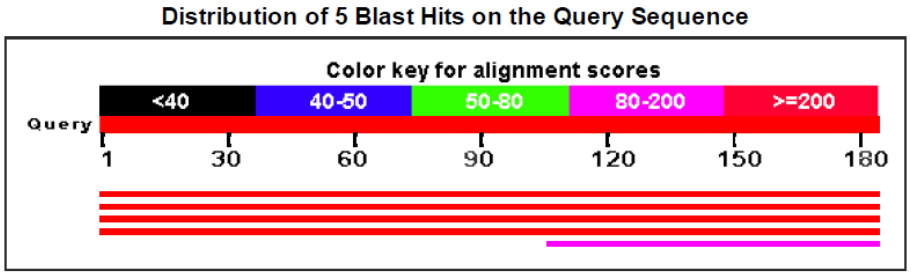
The locomotory rhythms of *F.palmarum* entrained to LD cycles and under constant conditions the rhythms showed a free running pattern with endogenous periodicity as reported earlier by Mammen and Jagota, 2011. The components of the molecular clock, namely *Per1,Per2,Per3, Cry 1, Cry2 and Bmal1* is expressed in the SCN of *F. palmarum.* The levels of these transcript showed significant variation between ZT0 and ZT12 (onset and offset of the entraining cue) with *Per1,Per2,Per3, Cry1* and *Cry2* levels significantly high during ZT12. The expression of *Bmal1* is high at ZT0 confirming to the phase difference between *Per2* and *Bmal1* transcript expression due to the immediate repressive effect of REVERBα on the *Bmal1* promoter.

C28. *Per1* in TA plasmid as template for sequencing. Primer M13F. Forward and Reverse Gene specific primers highlighted



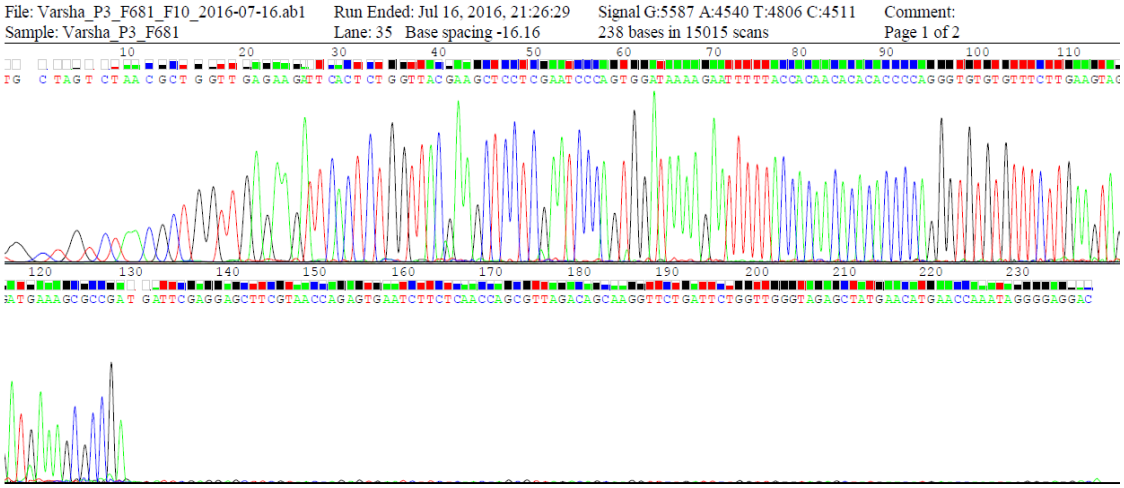
NCBI BLAST Report

Graphic Summary



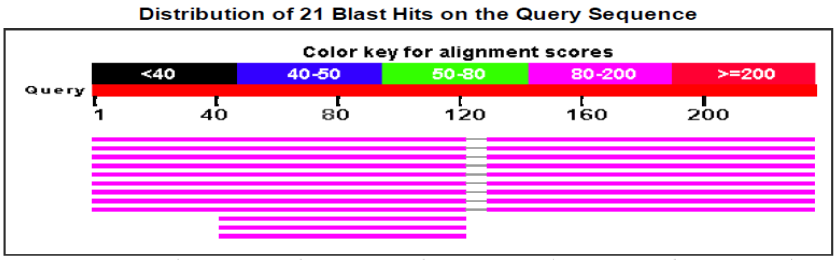
Description	Max score	Total score	Query cover	E value	Ident	Accession
Transcripts						
PREDICTED: Mus musculus period circadian clock 1 (Per1), transcript variant X2, mRNA	244	244	100%	9e-63	91%	XM_006532481.3
PREDICTED: Mus musculus period circadian clock 1 (Per1), transcript variant X1, mRNA	244	244	100%	9e-63	91%	XM_006532480.3
Mus musculus period circadian clock 1 (Per1), transcript variant 2, mRNA	244	244	100%	9e-63	91%	NM_001159367.1
Mus musculus period circadian clock 1 (Per1), transcript variant 1, mRNA	244	244	100%	9e-63	91%	NM_011065.4

C29. *Per3* PCR gel eluate as template for sequencing. Primer F681



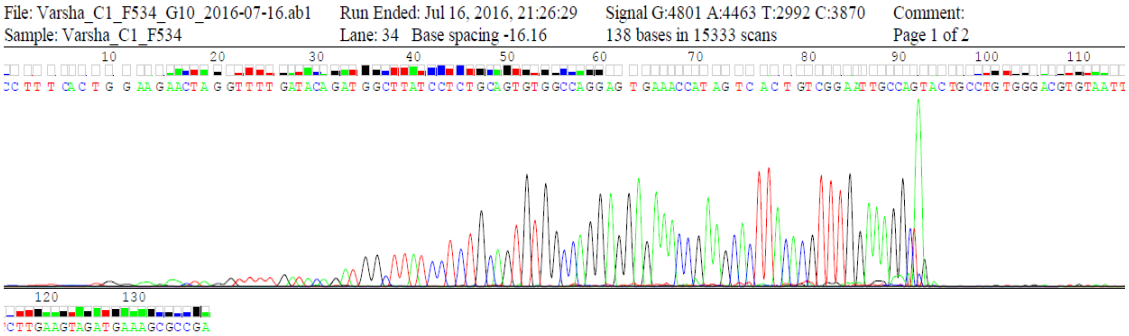
NCBI BLAST Report

Graphic Summary



Description	Max score	Total score	Query cover	E value	Ident	Accession
Transcripts						
PREDICTED: Mus musculus period circadian clock 3 (Per3), transcript variant X8, misc_RNA	187	347	96%	2e-45	94%	XR_001784108.1
PREDICTED: Mus musculus period circadian clock 3 (Per3), transcript variant X5, mRNA	187	347	96%	2e-45	94%	XM_017320034.1
PREDICTED: Mus musculus period circadian clock 3 (Per3), transcript variant X4, mRNA	187	347	96%	2e-45	94%	XM_011250203.2
PREDICTED: Mus musculus period circadian clock 3 (Per3), transcript variant X3, mRNA	187	347	96%	2e-45	94%	XM_017320033.1
PREDICTED: Mus musculus period circadian clock 3 (Per3), transcript variant X2, mRNA	187	347	96%	2e-45	94%	XM_017320032.1
PREDICTED: Mus musculus period circadian clock 3 (Per3), transcript variant X1, mRNA	187	347	96%	2e-45	94%	XM_017320031.1
Mus musculus period circadian clock 3 (Per3), transcript variant 3, mRNA	187	347	96%	2e-45	94%	NM_001289878.1
Mus musculus period circadian clock 3 (Per3), transcript variant 2, mRNA	187	347	96%	2e-45	94%	NM_011067.3
Mus musculus period circadian clock 3 (Per3), transcript variant 1, mRNA	187	347	96%	2e-45	94%	NM_001289877.1
PREDICTED: Mus musculus period circadian clock 3 (Per3), transcript variant X6, mRNA	128	128	34%	1e-27	95%	XM_017320035.1

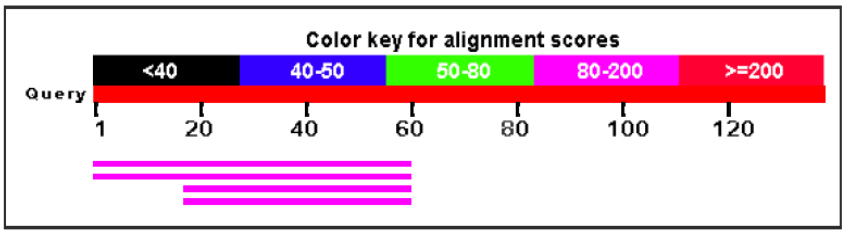
C29. *Cry1* PCR gel eluate as template for sequencing. Primer F534



NCBI BLAST Report

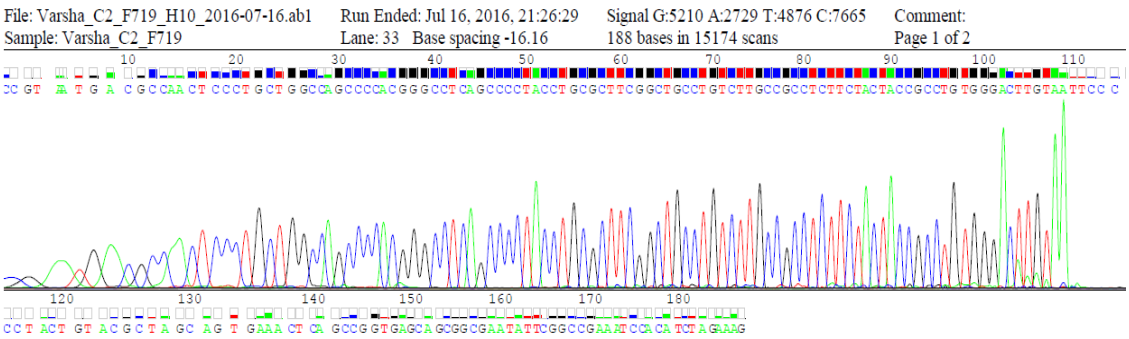
Graphic Summary

Distribution of 4 Blast Hits on the Query Sequence



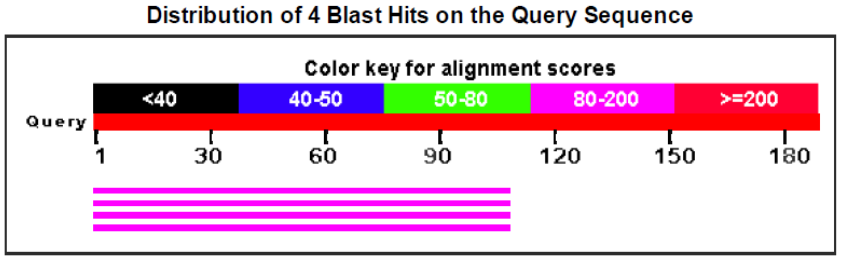
Description	Max score	Total score	Query cover	E value	Ident	Accession
Transcripts						
PREDICTED: Homo sapiens cryptochrome circadian clock 1 (CRY1), transcript variant X1, mRNA	100	100	43%	3e-19	97%	XM_017018832.1
Homo sapiens cryptochrome circadian clock 1 (CRY1), mRNA	100	100	43%	3e-19	97%	NM_004075.4

C30. *Cry2* PCR gel eluate as template for sequencing. Primer F719



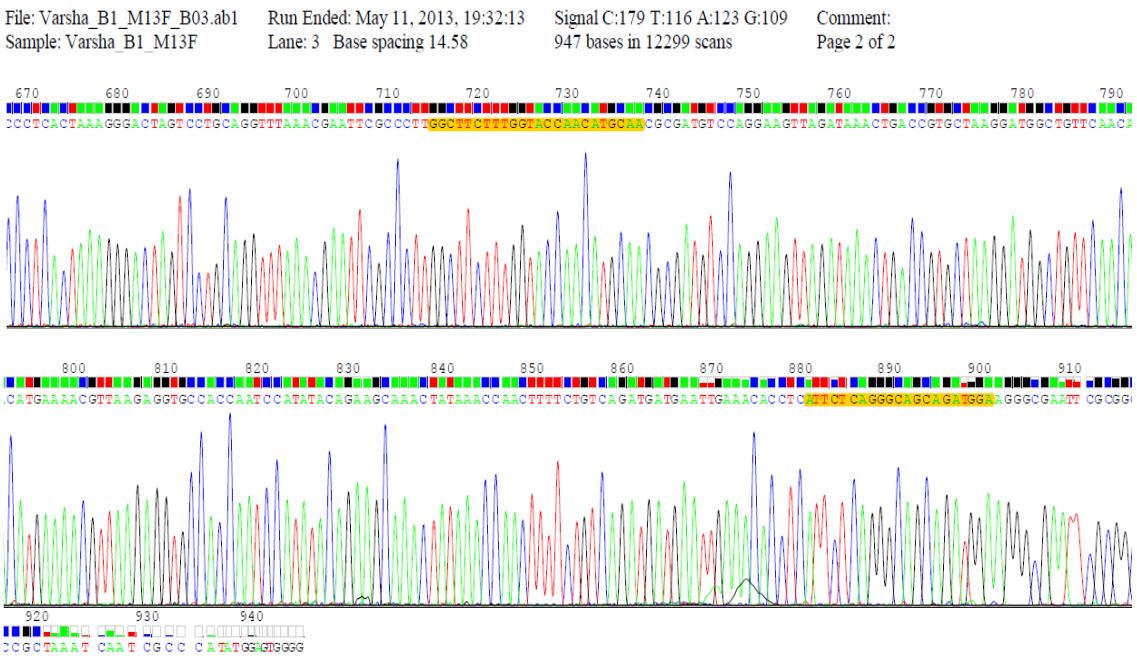
NCBI BLAST Report

Graphic Summary



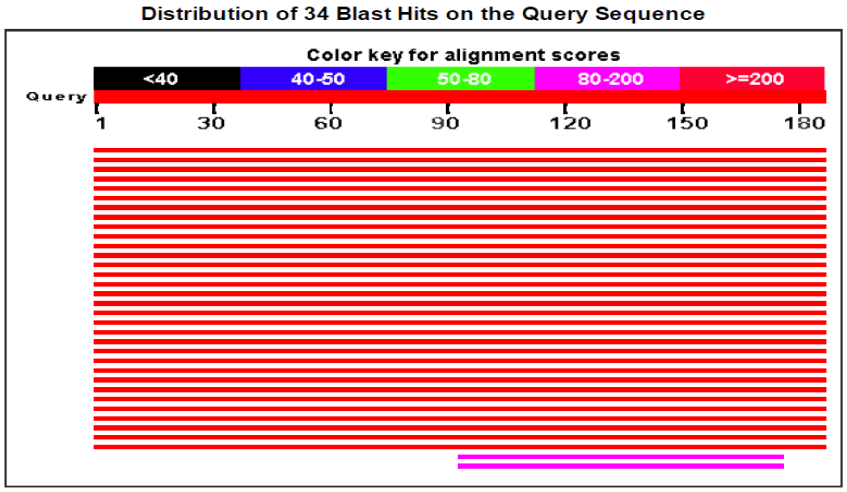
Description	Max score	Total score	Query cover	E value	Ident	Accession
Transcripts						
Homo sapiens cryptochrome circadian clock 2 (CRY2), transcript variant 2, mRNA	145	145	57%	2e-32	91%	NM_001127457.2
Homo sapiens cryptochrome circadian clock 2 (CRY2), transcript variant 1, mRNA	145	145	57%	2e-32	91%	NM_021117.3

C31. *Bmal1* -TA plasmid as template for sequencing. Primer M13 F. Forward and Reverse Gene specific primers highlighted.



NCBI BLAST Report

Graphic Summary



Description	Max score	Total score	Query cover	E value	Ident	Accession
Transcripts						
PREDICTED: Homo sapiens aryl hydrocarbon receptor nuclear translocator like (ARNTL), transcript variant X26, misc_RNA	291	291	100%	3e-76	95%	XR_001747880.1
PREDICTED: Homo sapiens aryl hydrocarbon receptor nuclear translocator like (ARNTL), transcript variant X25, misc_RNA	291	291	100%	3e-76	95%	XR_001747879.1
PREDICTED: Homo sapiens aryl hydrocarbon receptor nuclear translocator like (ARNTL), transcript variant X24, misc_RNA	291	291	100%	3e-76	95%	XR_001747878.1
PREDICTED: Homo sapiens aryl hydrocarbon receptor nuclear translocator like (ARNTL), transcript variant X23, misc_RNA	291	291	100%	3e-76	95%	XR_001747877.1
PREDICTED: Homo sapiens aryl hydrocarbon receptor nuclear translocator like (ARNTL), transcript variant X22, misc_RNA	291	291	100%	3e-76	95%	XR_001747876.1
PREDICTED: Homo sapiens aryl hydrocarbon receptor nuclear translocator like (ARNTL), transcript variant X21, mRNA	291	291	100%	3e-76	95%	XM_017017748.1
PREDICTED: Homo sapiens aryl hydrocarbon receptor nuclear translocator like (ARNTL), transcript variant X19, mRNA	291	291	100%	3e-76	95%	XM_017017747.1
PREDICTED: Homo sapiens aryl hydrocarbon receptor nuclear translocator like (ARNTL), transcript variant X18, mRNA	291	291	100%	3e-76	95%	XM_017017746.1
PREDICTED: Homo sapiens aryl hydrocarbon receptor nuclear translocator like (ARNTL), transcript variant X17, mRNA	291	291	100%	3e-76	95%	XM_011520112.2
PREDICTED: Homo sapiens aryl hydrocarbon receptor nuclear translocator like (ARNTL), transcript variant X16, mRNA	291	291	100%	3e-76	95%	XM_017017745.1

4. Characterization of the 1st intronic sequence of *F. palmarum Per2* a site for potential regulatory elements & *Per2* as a possible phylogenetic marker

The 5'UTR region of *Mus musculus* and *Rattus* appeared to be highly conserved. As the phylogenenetic utility of the *F.palmarum* 5'UTR has been explored in insects, we tried to compare the 5'UTR of rodents belonging to different suborders. Apart from *F. palmarum* (Scuridae), the 5'UTR of *Mesocricetus* (Muroidea) and *Cavia* (Caviomorpha*)* was elucidated by gDNA extraction. PCR analysis and sequencing for the comparative study.

PCR using forward E'E box primer and reverse primer within 1st intron of *F.palmarum* gDNA

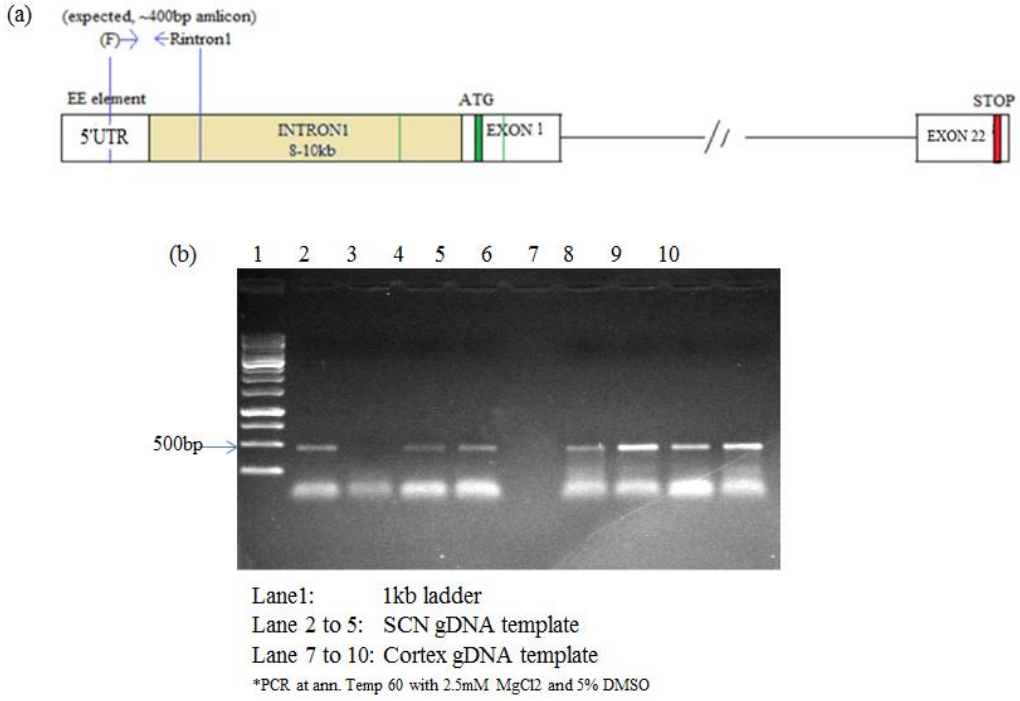


Fig.34. *Per2* gene of *F.palmarum* depicting the primer region. (b) The PCR product of the primer combination F cone/Rintron1 from both SCN and cortex gDNA (C28)

As there is no genomic information about the organism, we have compared the sequence with the previously characterized FconE/R113 report inorder to confirm the identity of the amplicon.

Genomic DNA extraction from Hamster and Guinea pig SCN and PCR with F conE/R106 primer.

The primer designed across the 1st intronic region of the *Per2* gene would generate an 8- 10 kb amplicon as predicted by the exonic architecture of *Per2* (Fig.18) The presence of a 5'UTR E'E box regulatory element is imperative for product formation. SCN gDNA isolated from hamster and guinea pig (Fig 35a) was used as template for the long range PCR reaction. The primer combination yielded PCR amplicons within the 8-10 and 6-8kb range with hamster and guinea pig gDNA , respectively. (Fig. 35 b,c). The amplicons were pooled for gel extraction and sequencing. (Fig.35 d,e).

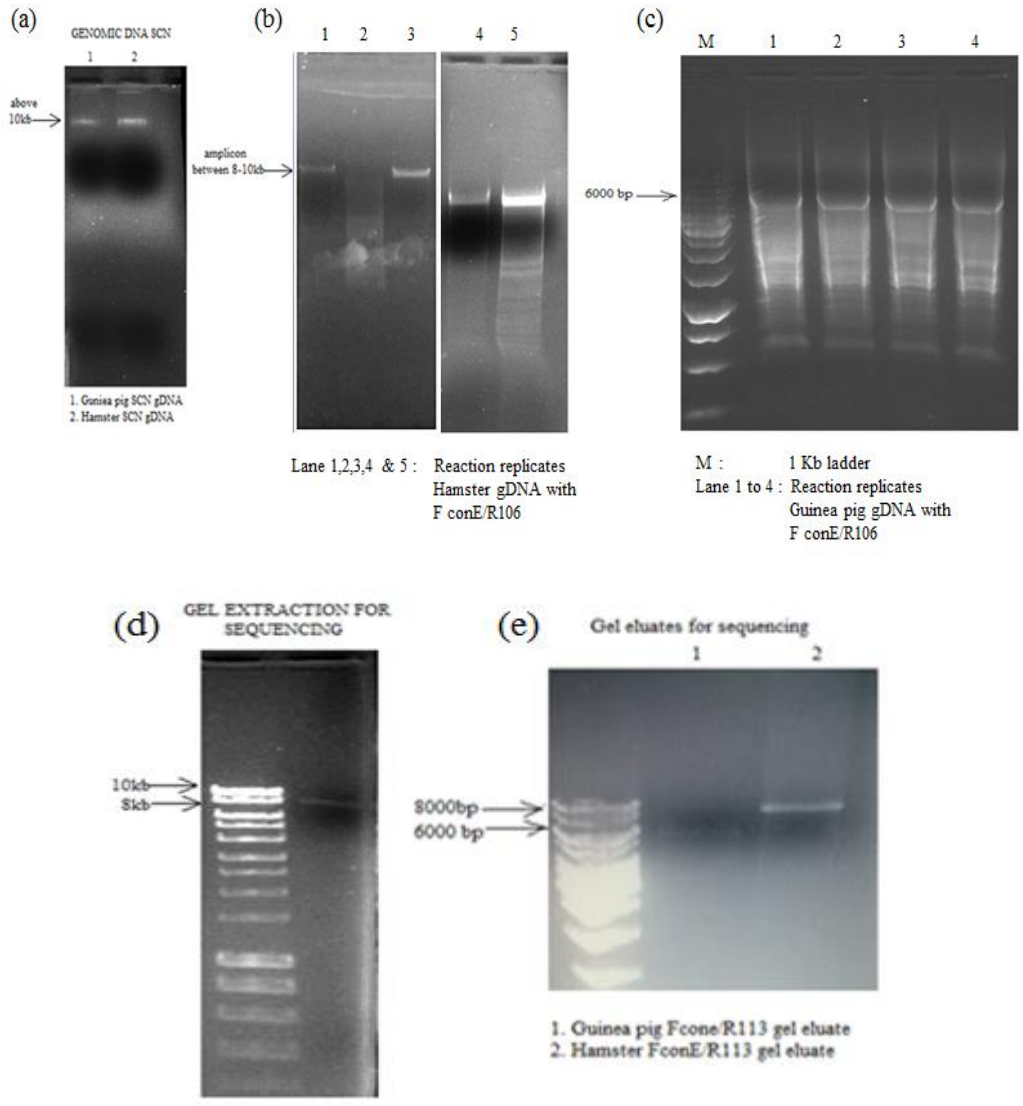
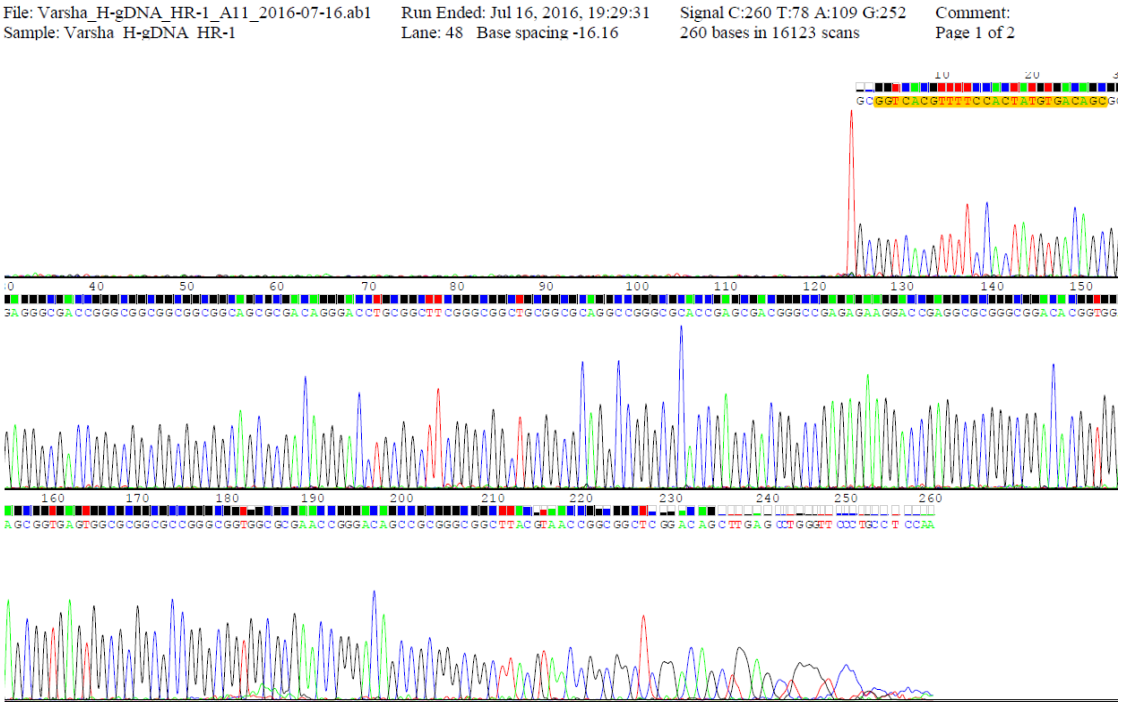


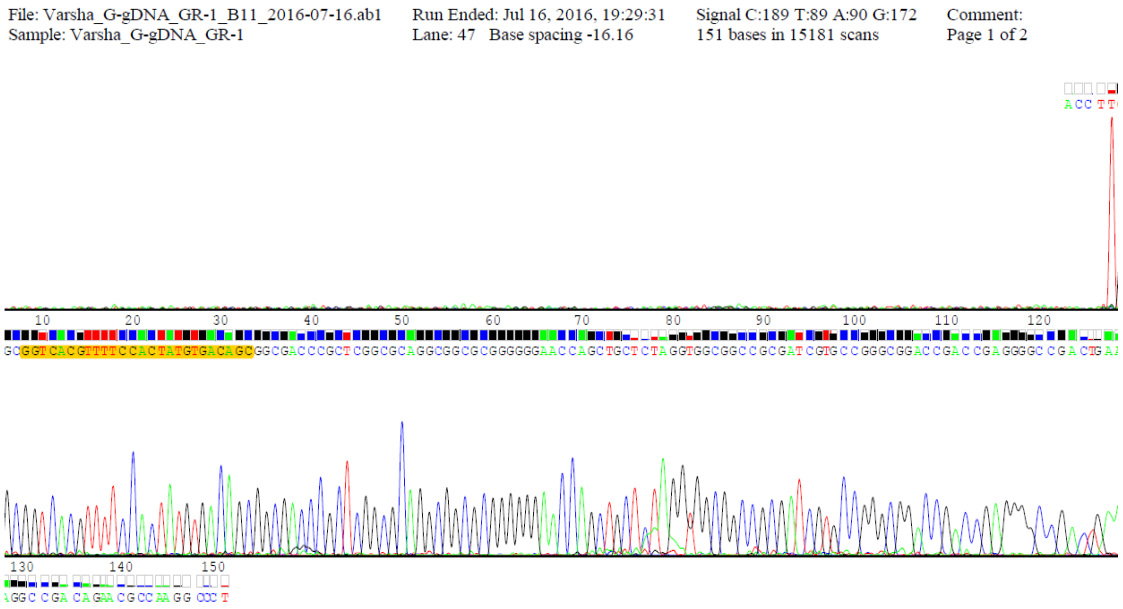
Fig.35.(a) gDNA extracted from the SCN of Guinea pig and Hamster (b) PCR products with Hamster gDNA sample (c) PCR products with Guinea pig sample. (d) gel extracted amplicon from hamster.(e) gel extracted amplicon from hamster and guinea pig gDNA reaction. The size difference of the *Per2* intergenic sequence in these animals is reflected in the amplicon size.

Sequence analysis confirms the presence of the E'E regulatory element in both hamster and guinea pig *Period 2* gene (C31, C32) indicating the possible usage of this highly conserved element in regulating rhythmic oscillations.

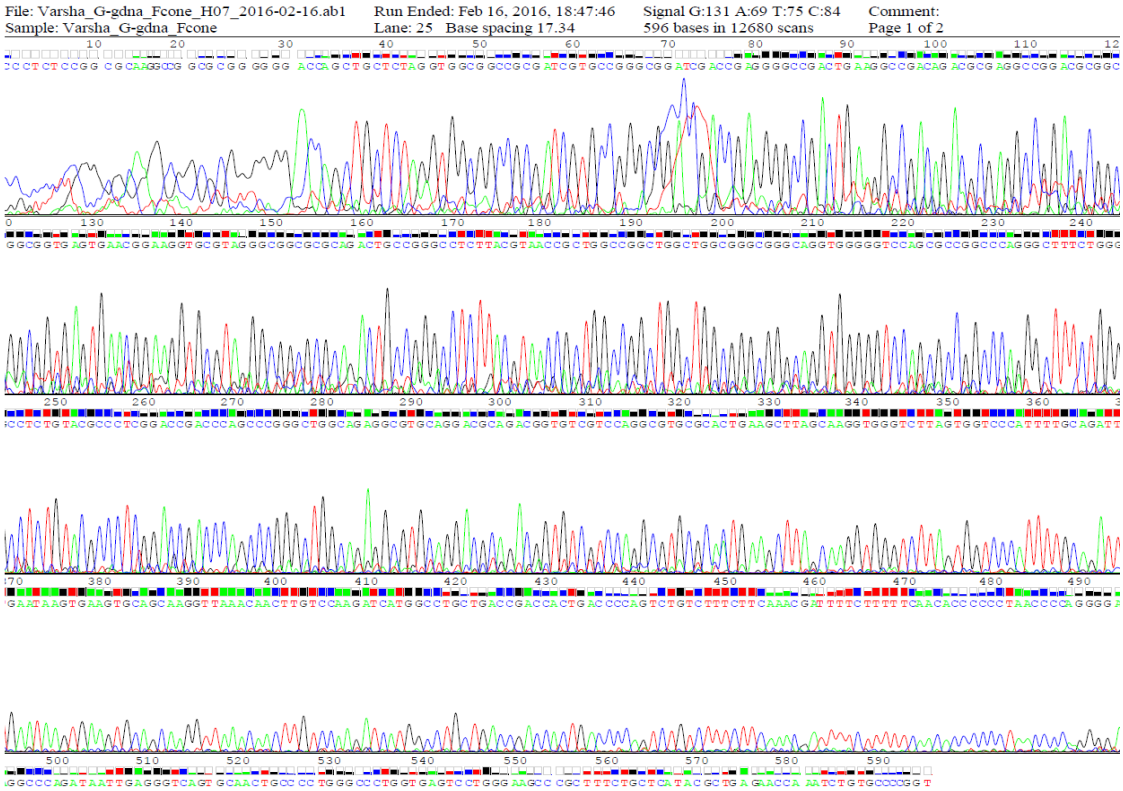
C31. Hamster SCN gDNA FconE/R106 PCR gel eluate as template.
Primer- Ham ConE +R1. (reverse primer downstream of E'E element) The E'E box is highlighted.



C32. Guinea pig SCN gDNA FconE/R106 PCR gel eluate as template.
Primer- Gui ConE +R1. The E'E box is highlighted.

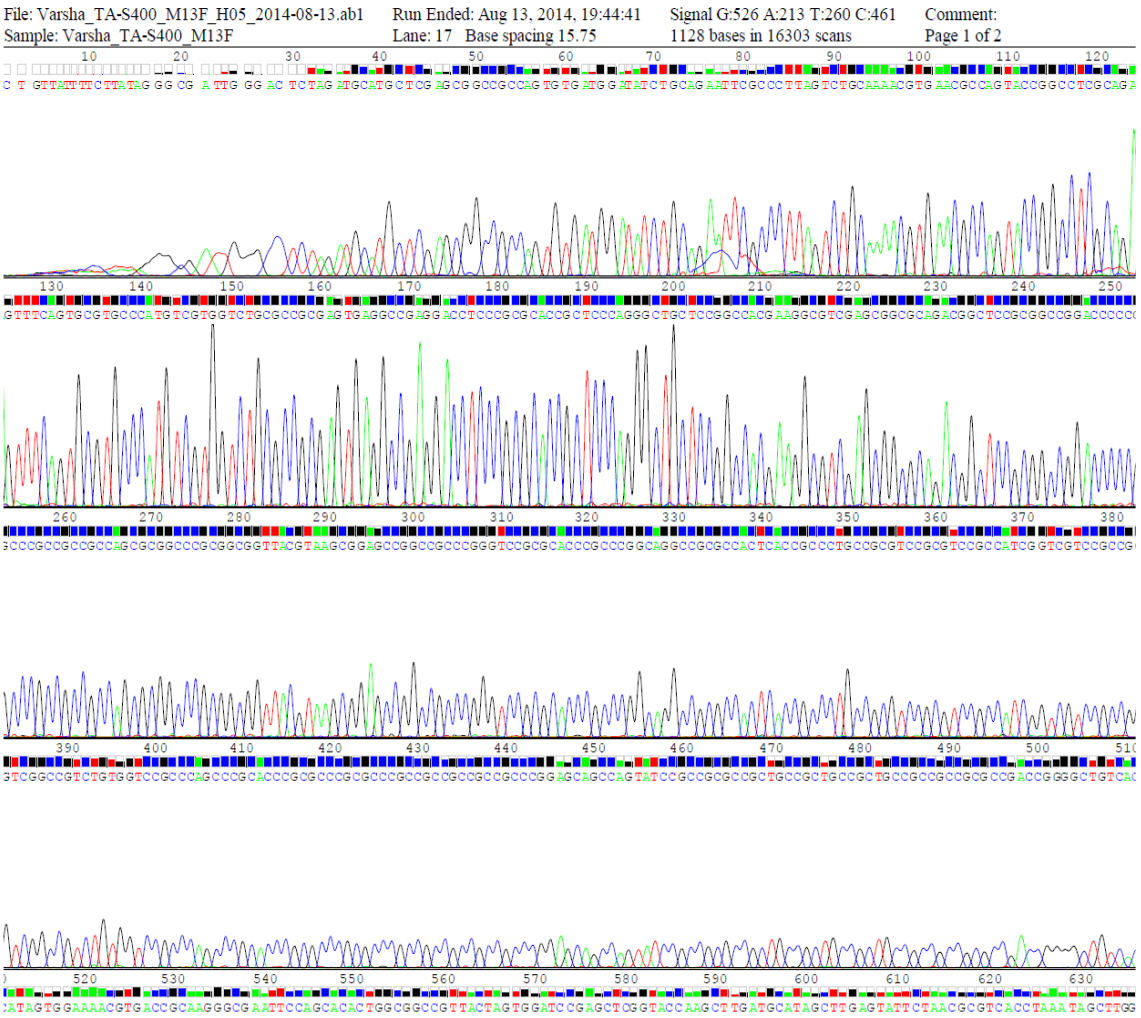


C.33. Guinea pig *Per2* 5'UTR. Primer FconE



C34. Hamster *Per2* 5'UTR. Primer FconE

C 35. *F. palmarum* 5'regualtory sequence (E'E box) in TA plasmid. Primer M13F

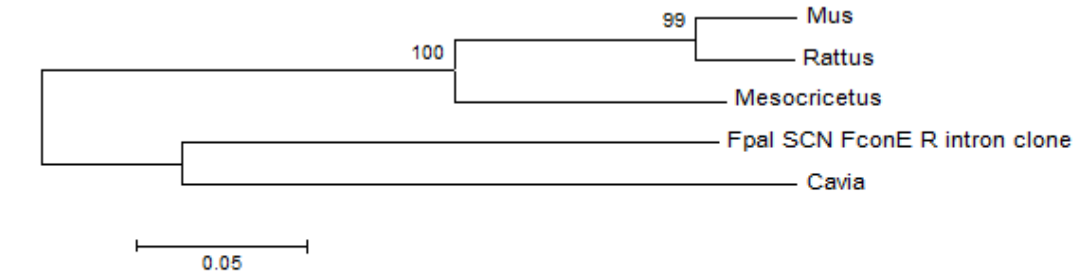


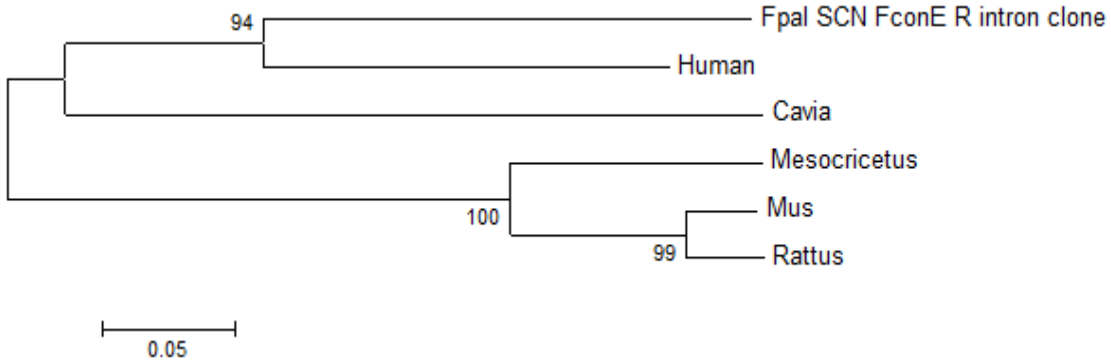
A6. Alignment of 4 sequencing reactions of *F. palmarum* EE box and 1st intronic sequence fpcon1,and 2 are PCR eluates, c400 and s400 (C35) are TA plasmids carrying the fragment amplified from the *F. palmarum* cortex and SCN, respectively. Highlighted in yellow-E 'E box and the reverse primer respectively.

```
                    *         20         *         40
fpcone1  :  ----------------------------------------  :   -
fpcone2  :  ----------------------------------------  :   -
c400     :  GCGGTCACGTTTTCCACTATGTGACAGCCCCGGGCGGCGCG  :  41
s400     :  GCGGTCACGTTTTCCACTATGTGACAGCCCCGGTCGGCGCG  :  41
            GCGGTCACGTTTTCCACTATGTGACAGCCCCGGGCGGCGCG

                    *         60         *         80
fpcone1  :  ----------------------------------------  :   -
fpcone2  :  ----------------------------------------  :   -
c400     :  GCGG------CGGCAGCGGCAGCGGCGCGGCGGATACTGGC  :  76
s400     :  GCGGCGGCAGCGGCAGCGGCAGCGGCGCGGCGGATACTGGC  :  82
            GCGGCGGCAGCGGCAGCGGCAGCGGCGCGGCGGATACTGGC

                    *        100         *        120
fpcone1  :  -----GCGGCGGCGGCGGCGGGAGCGGGAGCGGGTGGG  :  36
fpcone2  :  ----------------------------GGGCGCGGGTGCG  :  13
c400     :  TGCTCCGGGCGGCGGCGGCGGCGGGCGCGGGCGCGGGTGCG  : 117
s400     :  TGCTCCGGGCGGCGGCGGCGGCGGGCGCGGGCGCGGGTGCG  : 123
            TGCTCCGGGCGGCGGCGGCGGCGGGCGCGGGCGCGGGTGCG

                    *        140         *        160
fpcone1  :  GAGTAGGAGGAACAGACAATCTCAACCGGACGAATAGCGAT  :  77
fpcone2  :  GGCTGGGCGGACCACAGACGGCCGACCGGCGGACGACCGAT  :  54
c400     :  GGCTGGGCGGACCACAGACGGCCGACCGGCGGACGACCGAT  : 158
s400     :  GGCTGGGCGGACCACAGACGGCCGACCGGCGGACGACCGAT  : 164
            GGCTGGGCGGACCACAGACGGCCGACCGGCGGACGACCGAT

                    *        180         *        200
fpcone1  :  GGCGGACGCGGACGCGGGGTGAGGGGGAGGGGCGCGCCCGG  : 118
fpcone2  :  GGCGGACGCGGACGCGGCAGGGCGGTGAGTGGCGCGGCCTG  :  95
c400     :  GGCGGACGCGGACGCGGCAGGGCGGTGAGTGGCGCGGCCTG  : 199
s400     :  GGCGGACGCGGACGCGGCAGGGCGGTGAGTGGCGCGGCCTG  : 205
            GGCGGACGCGGACGCGGCAGGGCGGTGAGTGGCGCGGCCTG

                    *        220         *        240
fpcone1  :  CCGTGCGCGGGCGCGGACGCGGGGGTCCGGCTCCGCATACG  : 159
fpcone2  :  CCGGGCGGGTGCGCGGACCCGGGCGGCCGGCTCCGCTTACG  : 136
c400     :  CCGGGCGGGTGCGCGGACCCGGGCGGCCGGCTCCGCTTACG  : 240
s400     :  CCGGGCGGGTGCGCGGACCCGGGCGGCCGGCTCCGCTTACG  : 246
            CCGGGCGGGTGCGCGGACCCGGGCGGCCGGCTCCGCTTACG

                    *        260         *        280
fpcone1  :  CCACCGGCGGGGGTGGCGCTGGCGGGCGGGGGTCGGGGTCC  : 200
fpcone2  :  TAACCGCCGCGGGCCGCGCTGGCGGCGGCGGGCGGGGGTCC  : 177
c400     :  TAACCGCCGCGGGCCGCGCTGGCGGCGGCGGGCGGGGGTCC  : 281
s400     :  TAACCGCCGCGGGCCGCGCTGGCGGCGGCGGGCGGGGGTCC  : 287
            TAACCGCCGCGGGCCGCGCTGGCGGCGGCGGGCGGGGGTCC
```

```
                 *       300          *       320
fpcone1  :  GGACGCGGAGGCGTCTGCGCCGCTCTACGCGGTCGGGGCCT  : 241
fpcone2  :  GGCCGCGGAGCCGTCTGCGCCGCTCGACGCCTTCGTGGCCG  : 218
c400     :  GGCCGCGGAGCCGTCTGCGCCGCTCGACGCCTTCGTGGCCG  : 322
s400     :  GGCCGCGGAGCCGTCTGCGCCGCTCGACGCCTTCGTGGCCG  : 328
            GGCCGCGGAGCCGTCTGCGCCGCTCGACGCCTTCGTGGCCG

                *       340          *       360
fpcone1  :  GAGTGGCAGTGGTAGCCGGGAGCGCCTGGTCCTCAGTCTCG  : 282
fpcone2  :  GAGCAGCCCTGGGAGCGGTGCGCGGGAGGTCCTCGGCCTCA  : 259
c400     :  GAGCAGCCCTGGGAGCGGTGCGCGGGAGGTCCTCGGCCTCA  : 363
s400     :  GAGCAGCCCTGGGAGCGGTGCGCGGGAGGTCCTCGGCCTCA  : 369
            GAGCAGCCCTGGGAGCGGTGCGCGGGAGGTCCTCGGCCTCA

            *       380          *       400          *
fpcone1  :  GTCCAGACGACGACCTGTACATGCGCTCACACTGTAACACT  : 323
fpcone2  :  CTCGCGGCGCAGACCACGACATGGGCACGCACTGAAACTCT  : 300
c400     :  CTCGCGGCGCAGACCACGACATGGGCACGCACTGAAACTCT  : 404
s400     :  CTCGCGGCGCAGACCACGACATGGGCACGCACTGAAACTCT  : 410
            CTCGCGGCGCAGACCACGACATGGGCACGCACTGAAACTCT

                    420          *       440          *
fpcone1  :  GCGATGCCGGAATTGGCGTTCAGATTTTGCAGACAGA----  : 360
fpcone2  :  GCGAGGCCGGTACTGGCGTTCACGTTTTGCAGACCGAGAAA  : 341
c400     :  GCGAGGCCGGTACTGGCGTTCACGTTTTGCAGACT------  : 439
s400     :  GCGAGGCCGGTACTGGCGTTCACGTTTTGCAGACT------  : 445
            GCGAGGCCGGTACTGGCGTTCACGTTTTGCAGACTGAGAAA

                  460          *       480          *
fpcone1  :  ----------------------------------------  :   -
fpcone2  :  CAGGTTCAGAGAGGCCGCGCAACTTGTCTAAGGTCACAGAG  : 382
c400     :  ----------------------------------------  :   -
s400     :  ----------------------------------------  :   -
            CAGGTTCAGAGAGGCCGCGCAACTTGTCTAAGGTCACAGAG

                 500          *       520          *
fpcone1  :  ----------------------------------------  :   -
fpcone2  :  CCACTGAGTGACTTCTACACCAGACTGTGGCTCTTTAAACG  : 423
c400     :  ----------------------------------------  :   -
s400     :  ----------------------------------------  :   -
            CCACTGAGTGACTTCTACACCAGACTGTGGCTCTTTAAACG
```

The sequence information from this short stretch of nucleotides were used to generate a phylogenetic dendrogram (MEGA 6) also including *Mus musculus* , *Rattus norvegicus* and *Homo sapiens* data derived from NCBI





The similarity of the 5'UTR and partial intronic region of *Per2* confines to the phylogenetic organization of rodents, with Mus and Rattus showing most similarity followed by hamster (suborder Myodonta) thus are grouped together in the dendrogram. The *F. palmarum* (Scuridae) region forms a distinct group along with Cavia (Caviomorpha) as the region is more comparable between them. The use of an 'outgroup' i.e Primate sequence for phylogenetic analysis disapproves the utility of this region as a 'phylogenetic marker' in rodents. Notably, the *F. palmarum* sequence is more similar to the region in Homo sapiens (than to Cavia) which considering the regulatory role of this region could be an important observation

Apart from the partial 5'UTR region, the entire 10kb intronic region of *F. palmarum Per2* is being characterized by primer walking to know the presence of any intronic regulatory elements. The 5-to 6kb sequence that has been derived is not complete for efficient Bioinformatic analysis and hence not included in the thesis.

DISCUSSION & SUMMARY

The primary goal of the objectives was the characterization of a principal cogwheel of the molecular clock, *Period 2* in the diurnal rodent *F. palmarum*. The primary structure of the transcript was deduced using a comparative and functional approach from the genomic DNA and mRNA isolated from the SCN of this organism. The *F.palmarum Per2* has an open reading frame of 3756 bp and codes for a protein with 1251 amino acids which shows the presence of conserved functional domains as, PAS, PAC and FASP motifs. One of the most strikingly differentiating features in the transcript from this rodent is the presence of non-universal start codon, GUG in place of ATG. Importantly, the nucleotides flanking this alternative translation initiation site (aTIS) resemble a' Kozak' consensus sequence (Kozak,1987) and are similar to the other mammalian 'ATG' initiated *Per2* transcripts. The presence of an A or G at positions +4, -3, -9 (traditional Kozak consensus) also, C and G at -6 and -7 positions respectively, is found to be extremely conserved around the *Per2* start codon. The preference of A at -3 and G at the +4 positions in our model reflects that of primates and humans. Apart from influencing the affinity of translation initiation factors (Pisarev et al., 2006) differences between the pattern of nucleotide bias around the initiation site roughly reflects the evolutionary relationships of species (Nakagawa et al., 2007). Non AUG mediated translational initiation though relatively rare in mammals, GUG is the second most (after CUG) commonly preferred aTIS in eukaryotic mRNA (Wegrzyn, 2008). The efficient usage of GUG initiation codon relies on three major factors as, the nucleotide context (described above), unstructured upstream elements and stable secondary structure formation of downstream nucleotides (Takahashi, 2005; Tikole and Sankararamakrishnan, 2012). Though we are yet to analyse secondary structure formation of the mRNA in these regions the presence of a good Kozak context strengthens our argument that the translation machinery can initiate protein synthesis from this codon, also possibly ruling out any 'leaky scanning' from downstream AUGs (Kozak, 2002). Non AUG initiation codons are often used in addition to a downstream and in-frame AUG as a mechanism to generate functional diversity in polypeptides (Touriol, 2003) The GTG and ATG initiated isoforms of the Serpin protease inhibitors, 2B2 and 2B1 respectively is on such example of translational regulation in mammals (Hwang, 2005). Conversely, very few mRNAs are exclusively translated from a non AUG triplet and majority of them generate protein products that have cell signalling/regulatory functions (Tikole and Sankararamakrishnan, 2012). They include mRNAs of transcriptional enhancer factor (TEF1), rat bZIP transcription factor HLF36, human and mouse retinoic acid receptor β4, human PRP3, Drosophila erect wing, Arabidopsis AGAMOUS etc. (Takahashi et al., 2005). Most of the reported non AUG initiaton codon are not conserved in orthologues and paralogues as is the case presently observed in *Period2* and its paralogues. The presence of 'GTG' in our model could be a result of sporadic point mutation of the AUG triplet that took place after the divergence of paralogues and orthologues and which was also sustained due to the presence of an appropriate 'translation context' (Takahashi et al.,2005) that did not hinder translation. The presence of GUG as a highly conserved start codon across various taxa is reported by the same group for eukaryotic translation initiation factor, NAT1. Ambiguities in codon usage among different species were reported in yeast by Kawaguchi et al in1989.

CUG, a universal leucine codon can be translated as Serine in some species owing to a point mutation on tRNA and the choice of leucine or serine was found to be random.

In prokaryotes apart from AUG, GUG and UUG act as start codons. Inspite of alternative usage of initiator triplets all of them are translated as 'Methionine' by initiator tRNA (Lobanov et al., 2010). Even in Eukaryotes addition of 'Methionine' at non cognate codons may result from base pairing of Met-tRNA to anticodons that are not perfectly complementary, as most of the non canonical initiators differ from AUG by only a single base pair. This hypothesis was demonstrated by the addition of Methionine residue in dihydrofolate reductase mRNA even when the initiator AUG was mutated to ACG, CUG, GUG, UUG, AUA, AUC or AUU (Peabody, 1989). Similarly, incorporation of Methionine at non AUG translation initiation site has been reported *in-vitro* for human transcription factor SP3 (Hernandez et al., 2002) and also in *C.albicans* P2A protein (Abramczyk et al., 2003). However, using mutant initiator tRNAs initiation of protein synthesis in COS1 cells was possible with both methionine and valine (Drabkin and Rajbhandary,1998) The authors also report the relative stability and efficiency of Valine (GUG) over other aminoacids as an initiator of protein synthesis. Valyl-tRNA and isoleucyl-tRNA have most stable linkages between an amino acid and tRNA (Matthaei et al., 1966). A dual translational initiation in response to aminoacid deprivation is reported in c-myc1 protein (Hann, 1995) In Major Histocompatibility Complex (MHC) class I bound peptides the CUG initiation codon is decoded by Leucine instead of Methionine proving that the eIF2-GTP-methionine tRNA may not be the only translational ternary complex (Touriol et al., 2003). The variations in usage of the genetic languages observed across cellular organelles and different taxonomic groups contradicts the hypothesis of a 'frozen accident' proposed by Francis Crick which assumes that all the present day organisms use the universal and invariant genetic code (Lobanov et al., 2010). With respect to the present observation in *F. palmarum* experimental validation by N-terminal sequencing of PER2 in parallel with *in-vitro* expression of the coding region in eukaryotic cell lines are inevitable to understand the scenario of translation initiation and regulation in this species. Though cap-independent translation mediated via IRES is known in *Per1* (Lee et al., 2011), non-AUG initiated translation is never reported for any clock gene to date.

The presence of a transcript variant of *Per2*, named as *FpPer2v* was isolated from the SCN of our model. *Per2v* is distinct from the basic transcript in the presence of an extra 503 bp insertion between exons 9 and 10 which introduces a Premature termination codon (PTC) in the transcript possibly synthesizing a truncated protein product of 454 amino acids that also retains one of the PAS domains. Based on the exonic architecture of *Per2* that was constructed for the present study it can be reasonably assumed as an event of Intronic Retention (IR). Moreover a subsequent deletion of exons 13 to 18 was characteristic of the variant, the significance of this deletion after the PTC may be less important. However, after the PTC and couple of in

frame STOP codons an ORF coding for 367 amino acids with C terminus PAC domain is also observed in the variant. Strikingly the the FASP motif is absent in the PER2v. A splice variant of *hPer2*, named as *Per2s* (Per2 short) annotated in GeneBank with the accession number AB012614 was recently detected in human keratinocyte cell line. *Per2s* has 1215 bp and codes for a protein of 404 amino acids (Avitabile et al., 2013) and is the only splice variant reported for *Per2*. We could map the deletion events that results in *Per2s* to E9, the skipping of this exon results in alteration of the reading frame leading to premature termination of the transcript soon after translation of E10. The FASP motif is also absent in PER2s and similar to the *F. palmarum* PER2v the co-activator-like protein/protein interaction motif (LXXLL, 308–313 AA) of PER2, which mediates the binding to several nuclear receptors including REV-ERBα and PPARα is conserved in PER2s. The presence of these variants may be indicative of highly 'active splicing' events at E8/E9 leading to two distinct consequences, exon skipping (*Per2s in humans*) or intronic retention (as in *Per2v*) culminating in a shorter *Per2* transcript. *Per2s* has a circadian profile of expression, the protein is localized to the nucleolus and is a crucial messenger of genomic stress by altering the phase properties of core clock components (Avitabile et al., 2013). It is of our interest to detect the protein product of *Per2v* inorder to truly understand the functional significance of our observation. Regulated Unproductive Splicing and Translation (RUST) that involves the addition of PTC by Alternate splicing events mainly by Intronic Retention , Exon skipping or insertion of an Alternative Poison Casette Exon (PCE) is widely prevalent in Arabidopsis clock genes. The abundance of a full-length protein coding transcript can be regulated by shifting the pre mRNA splicing towards PTC-containing isoforms that are degraded by the NMD pathway. It also results in truncated protein products that are believed to alter the functioning of its normal full length counterpart. Acting at the posttranscriptional level RUST acts as an immediate response to the variations in the external environment (Filichkin and Mockler, 2012). The role of AS mechanism involving intronic retention and exon skipping is well studied in neuronal differentiation and regeneration (Boutz et al., 2007; Xu et al., 2008) however its significance in mammalian clock gene regulation is yet to be understood. Though the presence of such variants with insertions/deletions is reported in *Per3* of a subterranean mole rat (Avivi et al., 2002) its functional significance is not yet explored.

The transient nature of photic induction is attributed to the difference in the turn over rate of Per2/PER2 in comparison to the sustained endogenous expression of the transcript as well as its protein product (Shigeyoshi et al., 2007; Field et al., 2000). CRY1 is known to mediate PER2 stability by favoring phosphorylation at the FASP motifs (Vanselow et al., 2007) hence the non photo inducible CRY could be a limiting factor that affects the half-life of the PER2 protein subsequent to a photic impulse. The fact that photic induction of *Per* happens via the CRE elements (not via E box) also give room for speculating the presence of a variant that would solely communicate the photic information for clock resetting, similar to the nucleolar expressed *Per2s* devoid of a FASP binding domain which acts as a messenger of cellular stress (Avitabile et al., 2013). Considering the functional heterogeneity of SCN with its photically inducible

'core' conveying light information to the endogenous 'shell' neurons, localization of the deduced transcripts of *Per2* in the context of the 'core' marker *Vip* (Mammen and Jagota, 2011) would throw some light on the validity of this hypothesis. For specific cRNA probe generation the nucleotide sequence of the major coupling agent of SCN, *Vip* was cloned and sequenced from the master clock of our animal model. The results proved the advanced phylogenetic position of *F. palmarum*, as the transcript was more comparable to that of primate sequences. The advantage of phylogenetically advanced animal models in the study of neurodegenerative and behavioural disorders is exemplified by the diurnal rodent *O. degus.* Similar observations emphasizes the emerging significance of our model in the study of chronobiology.

The utility of this animal model in chronobiology research is further extended with the detection and identification of other Period paralogues, *Per1* and *Per3* as well as *Cry1*, *Cry2* and *Bmal1* in the SCN. These transcripts showed a circadian pattern of expression under entrainable conditions with significant variation in expression profile between ZT0 and ZT12 (onset and offset of the entraining cue) *Per1,Per2, Per3, Cry1* and *Cry2* levels were high during ZT12 whereas The expression of *Bmal1* was high at ZT0. This phase difference in *Per2* and *Bmal1* expression is associated with the immediate feedback repressive effect of REVERBα on the *Bmal1* promoter unlike PER2 which has to mature as a repressor via progressive and site specific PTMs as reported by Ripperger and Albrecht, 2012.

The study also confirms the presence of a highly conserved regulatory element, the E'E box in the 5'UTR of *F. palmarum, Per2.* The region was also detected in hamster as well as the guinea pig gene. Among the transcriptional controls in mammals, 3 major clock components namely the E-box, RORE, and DBPE are significant in robust rhythm generation. However, studies by Nakahata and group (2008) proved that apart from a single E box closely spaced 2 E box- like elements (EE, E'E or E2) is important for circadian oscillation of clock gene and clock controlled genes. Dimeric complexes of transcription factors such as the bHLH family transcription factor HIF1 or the bZip family transcription factor E4BP4 is believed to bind to E2. Also, different forms of CLOCK and BMAL1 complex may also associate with E2 mediated transcription, further modulating the core loop of the TTLF model (Nakahata et al., 2008). Exonic architecture of the *Per2* gene was relatively conserved in rat, mouse and humans with the nucleotide number of individual exons and introns very much comparable between them. The 1st intronic sequence was observed to be the largest, 8 to 10 kb in mouse and human respectively. The *F. palmarum* intergenic sequence was around 10kb while it was approximately 8kb in hamster and 6 kb in guinea pig. Phylogenetic analysis of the 5'UTR and partial intergenic sequence of *Per2* indicated more similarity in this region between *F. palmarum* and *Homo sapiens*. It is important to note here that non coding sequences that are conserved between genomes of distantly related Taxa are considered to have specific functionality than sequences that are unique to a particular genome (Weinreb, 2001). The similarity in this regulatory region could hence be of particular significance with respect to diurnality. However, extensive

comparative analysis involving diverse animal groups would be important to come to any definitive conclusion. With the possibility of yet to be identified promoter elements contributing to circadian rhythm generation characterization of the largest intronic element of *Per2* (Intron1) is being carried out as a part of this work. It is also reported that first introns have blocks of highly conserved sequences that are enriched for several chromatin marks indicative of active regulatory regions (Park et al., 2014). However, at the moment only partial information could be generated in this regard that is insufficient for any reliable Bioinformatic analysis.

SUMMARY

As a part of characterizing the molecular components of the biological clock in the diurnal rodent, *F. palmarum* the present work focuses on cloning and characterization of *Period 2* gene from the SCN. The transcript codes for a protein of 1251 AA, and similar to other mammalian PER2 has conserved PAS, PAC and FASP structural motifs. Distinctly, the transcript shows the presence of 'GUG' as the initiator triplet in the context of a highly conserved Kozak consensus. The presence of a non AUG 'start' codon has never been reported for a clock gene and it indicates the non universality of genetic language across different animal species. A transcript variant, Per2v is also reported, novel post-transcriptional regulatory mechanisms in generating specific functional variants is speculated with this observation. In comparison to the newly reported Per2s in humans, the absence of a FASP motif in Per2s with the protein/protein interaction motif, LXXL intact a distinct functional goal is predicted for this variant. Localization of these transcripts within the functionally heterogenous SCN core and shell, using Vasoactive Intestinal peptide (*Vip*) as a marker for core neurons would be important to support our hypothesis. The structural similarity of *F.palmarum Vip* towards Human and Primate sequences indicate an advanced phylogenetic status of our model, similar to the more recent diurnal models in chronobiology. Also, the 5'UTR and partial intergenic sequences of Per2 in this rodent bears resemblance to that of Homo sapiens possibly indicating a comparable functional significance for the region. The Per2 5'UTR regulatory element EE box appeared to be highly conserved in *F.palmarum*, hamster and guinea pig pointing towards its functional significance in conjunction with the single E-Box element that drive endogenous rhythms. The characterization of the 1st intronic sequence of *Per2* in order to identify potential regulatory elements is also part of this work. This diurnal rodent can be an ideal model in circadian biology also with respect to the expression profile of major clock genes, *Per1, Per3, Cry1, Cry2* and *Bmal1* in the master clock.

This work is a significant step towards establishing a reliable diurnal animal model for the study of circadian clock which could be more comparable to that of humans. Though functional analysis of our findings are crucial for explaining species specific regulatory mechanisms, our model being phylogenetically more advanced and hence bearing more resemblance in many of its key molecular clock components to that of

higher primates it is of no doubt that the characterization of *F.palmarum* clock could offer better insights into the mechanisms of diurnality.

References

Abe, H., Honma, S., Shinohara, K and Honma, K.I (1999) Circadian modulation in photic induction of Fos-like immunoreactivity in the suprachiasmatic nucleus cells of diurnal chipmunk, *Eutamias asiaticus*. J Comp Physiol A. 176: 159-167.

Abramczyk, D., Tchórzewski, M and Grankowski, N (2003) Non-AUG translation initiation of mRNA encoding acidic ribosomal P2A protein in *Candida albicans*. Yeast. 20:1045-1052.

Ananthasubramaniam, B., Herzog, E.D and Herzel, H (2014) Timing of neuropeptide coupling determines synchrony and entrainment in the mammalian circadian clock. PLoS Comput Biol. 10: e1003565

Antle, M.C and Silver, R (2005) Orchestrating time: arrangements of the brain circadian clock. Trends Neurosci. 28: 145-151.

Ashkenazy-Frolinger, T., Einat, H and Kronfeld-Schor, N (2014) Diurnal rodents as an advantageous model for affective disorders: novel data from diurnal degu *(Octodon degus)*. J Neural Transm (Vienna). 122: S35-S45.

Avitabile, D., Genovese, L., Ponti, D., Ranieri, D., Raffa, S., Calogero, A and Torrisi, M.R (2013) Nucleolar localization and circadian regulation of Per2S, a novel splicing variant of the *Period 2* gene. Cell Mol Life Sci. 71:2547-2559.

Avivi, A., Oster, H., Joel, A., Beiles, A., Albrecht, U and Nevo, E (2002) Circadian genes in a blind subterranean mammal II: conservation and uniqueness of the three Period homologs in the blind subterranean mole rat, *Spalax ehrenbergi* superspecies. Proc Natl Acad Sci. 99:11718-11723.

Ban, Y., Shigeyoshi, Y and Okamura, H (1997) Development of vasoactive intestinal peptide mRNA rhythm in the rat suprachiasmatic nucleus. J Neurosci. 17: 3920-3931.
Barbosa, C., Peixeiro, I and Romão, L (2013) Gene expression regulation by upstream open reading frames and human disease. PLoS Genet. 9: e1003529.

Bauzer, L.G.S.R., Gesto, J.S.M., Souza, N.A., Ward, R.D., Hamilton, J.G.C., Kyriacou, C.P and Peixoto, A.A (2002) Molecular divergence in period gene between two putative sympatric species of *Lutzomyia longipalpis* Complex. Mol.Bio.Evol.19: 1624-1627.

Bell-Pedersen, D., Cassone, V.M., Earnest, D.J., Golden, S.S., Hardin, P.E., Thomas, T.L and Zoran, M.J (2005) Circadian rhythms from multiple oscillators: lessons from diverse organisms. Nat Rev Genet. 6: 544-556.

Boutz, P.L., Stoilov, P., Li, Q., Lin, C.H., Chawla, G., Ostrow, K., Shiue, L., Ares, M. Jr and Black, D.L (2007) A post-transcriptional regulatory switch in polypyrimidine

tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev. 21: 1636-1652.

Brancaccio, M., Enoki, R., Mazuski, C.N., Jones, J., Evans, J.A and Azzi, A (2014) Network-mediated encoding of circadian time: the suprachiasmatic nucleus (SCN) from genes to neurons to circuits, and back. J Neurosci. 34: 15192-15199.

Brunner, M., Simons, M.J and Merrow, M (2008) Lego clocks: Building a clock from parts. Genes Dev. 22: 1422-1426.

Cao, R., Gkogkas, C.G., de Zavalia, N., Blum, I.D., Yanagiya, A., Tsukumo, Y., Xu, H., Lee, C., Storch, K.F., Liu, A.C., Amir, S and Sonenberg, N (2015) Light-regulated translational control of circadian behavior by eIF4E phosphorylation. Nat Neurosci. 18: 855-862.

Carpen, J.D., Archer, S.N., Skene, D.J., Smits, M and 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: 293-297.

Chen, Z., Yoo, S.H., Park, Y.S., Kim, K.H., Wei, S., Buhr E., *et al.* (2012) Identification of diverse modulators of central and peripheral circadian clocks by high-throughput chemical screening. Proc Natl Acad Sci. 109: 101-106.

Cheng, H.Y., Papp, J.W., Varlamova, O., Dziema, H., Russell, B., Curfman, J.P., Nakazawa T., Shimizu, K., Okamura, H., Impey, S and Obrietan, K (2007) microRNA modulation of circadian-clock period and entrainment. Neuron. 54: 813-829.

Chiu, J.C., Ko, H.W and Edery, I (2011) NEMO/NLK phosphorylates PERIOD to initiate a time-delay phosphorylation circuit that sets circadian clock speed. Cell. 145: 357-370.

Chomczynski, P and Sacchi, N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem. 162:156-159.

Chong, S.Y., Ptacek, L.J and Fu, Y.H (2012) Genetic insights on sleep schedules: this time, it's PERsonal. Trends Genet. 28: 598-605.

Colwell, C.S (2011) Linking neural activity and molecular oscillations in the SCN. Nat Rev Neurosci. 12: 553-569.

Cowling, V.H (2010) Regulation of mRNA cap methylation. Biochem. J. 425: 295-302.

Coyne, J.A (1992) Genetics and speciation. Nature. 355: 511-515.

Daan, S., Albrecht, U., van der Horst, G.T., Illnerová, H., Roenneberg, T., Wehr, T.A and Schwartz, W.J (2001) Assembling a clock for all seasons: are there M and E oscillators in the genes? J Biol Rhythms. 16: 105-116.

David, R and Wedlich, D (2001) PCR-based RNA probes: a quick and sensitive method to improve whole mount embryo in situ hybridizations. Biotechniques. 30:769-72, 74.

Decoursey, P.J (2004) The behavioral ecology and evolution of biological timing systems. In: Dunlap JC, Loros JJ, DeCoursey PJ, (Ed.), Chronobiology: Biological timekeeping. Sunderland: Sinauer Associates. pp. 27-62.

de la Iglesia, H.O (2007) In situ hybridization of suprachiasmatic nucleus slices. Methods Mol Biol. 362:513-531.

de la Iglesia, H.O., Meyer, J., Carpino, A. Jr and Schwartz, W.J (2000) Antiphase oscillation of the left and right suprachiasmatic nuclei. Science. 290: 799-801.

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

Drabkin, H.J and RajBhandary, U.L (2008) Initiation of protein synthesis in mammalian cells with codons other than AUG and amino acids other than methionine. Mol Cell Biol. 18: 5140-5147.

Duffy, J.F and Czeisler, C.A (2000) Effect of Light on Human Circadian Physiology. Sleep Med Clin. 4: 165-177.

Field, M.D., Maywood, E.S., O'Brien, J.A., Weaver, D.R., Reppert, S.M and Hastings, M.H (2000) Analysis of clock proteins in mouse SCN demonstrates phylogenetic divergence of the circadian clockwork and resetting mechanisms. Neuron. 25: 437-447.

Filichkin, S.A and Mockler, T.C (2012) Unproductive alternative splicing and nonsense mRNAs: a widespread phenomenon among plant circadian clock genes. Biol Direct. 7:20.

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

Fu, L., Pelicano, H., Liu, J., Huang, P and Lee, C (2002) The circadian gene *Period2* plays an important role in tumor suppression and DNA damage response *in vivo*. Cell. 111: 41-50.

Fustin, J.M., Doi, M., Yamaguchi, Y., Hida, H., Nishimura, S., Yoshida, M., Isagawa, T., Morioka, M.S., Kakeya, H., Manabe, I and Okamura, H (2013) RNA-methylation-dependent RNA processing controls the speed of the circadian clock. Cell. 155: 793-806.

Gerkema, M.P., Davies, W.I., Foster, R.G., Menaker, M and Hut, R.A (2013) The nocturnal bottleneck and the evolution of activity patterns in mammals. Proc Biol Sci. 280: 20130508.

Golombek, D.A and Rosenstein, R.E (2010) Physiology of circadian entrainment. Physiol Rev. 90: 1063-1102.

Gustafson, C.L and Partch, C.L (2015) Emerging models for the molecular basis of mammalian circadian timing. Biochemistry. 54: 134-149.

Hamada, T., LeSauter, J., Venuti, J.M and Silver, R (2001) Expression of Period genes: rhythmic and nonrhythmic compartments of the suprachiasmatic nucleus pacemaker. J Neurosci. 21: 7742-7750.

Hann, S.R (1995) Methionine deprivation regulates the translation of functionally-distinct c-Myc proteins. Adv Exp Med Biol. 375:107-116.

He, B and Chen, Z (2016) Molecular Targets for Small-Molecule Modulators of Circadian Clocks. Curr Drug Metab. 17: 503-512.

Helfrich-Förster, C (2005) Neurobiology of the fruit fly's circadian clock. Genes Brain Behav. 4:65-76.

Hernandez, E.M., Johnson, A., Notario, V., Chen, A and Richert, J.R (2002) AUA as a translation initiation site *in vitro* for the human transcription factor Sp3. J Biochem Mol Biol. 35: 273-282.

Hirano, A., Yumimoto, K., Tsunematsu, R., Matsumoto, M., Oyama, M., Kozuka-Hata, H., Nakagawa, T., Lanjakornsiripan, D., Nakayama, K.I and Fukada, Y (2013) FBXL21 regulates oscillation of the circadian clock through ubiquitination and stabilization of cryptochromes. Cell. 152: 1106-1118.

Hwang, S.R., Garza, C.Z., Wegrzyn, J.L and Hook, V.Y (2005) Demonstration of GTG as an alternative initiation codon for the serpin endopin 2B-2. Biochem Biophys Res Commun. 327: 837-844.

Jagannath, A., Butler, R., Godinho, S.I., Couch, Y., Brown, L.A., Vasudevan, S.R., Flanagan, K.C., Anthony, D., Churchill, G.C., Wood, M.J., Steiner, G., Ebeling, M., Hossbach, M., Wettstein, J.G., Duffield, G.E., Gatti, S., Hankins, M.W., Foster, R.G and Peirson, S.N (2013) The CRTC1-SIK1 pathway regulates entrainment of the circadian clock. Cell. 154: 1100-1111.

Jagota, A (2006) Suprachiasmatic Nucleus: Centre for circadian time keeping system in mammals. Proc Indian Natl Sci Acad. 71: 285-300.

Jagota, A., de la Iglesia, H.O and Schwartz, W.J (2000) Morning and evening circadian oscillations in the suprachiasmatic nucleus *in vitro*. Nat Neurosci. 3: 372-376.

Jagota, A and Reddy, M.Y (2007) The effect of curcumin on ethanol induced changes in suprachiasmatic nucleus (SCN) and pineal. Cell Mol Neurobiol. 27: 997-1006.

Janich, P., Arpat, A.B., Castelo-Szekely, V., Lopes, M and Gatfield, D (2015) Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. Genome Res. 25:1848-1859.

Johnson, C.H., Mori, T and Xu, Y (2008) A cyanobacterial circadian clock work. Curr. Biol. 18: R816-R825.

Kawaguchi, Y., Honda, H., Taniguchi-Morimura, J and Iwasaki, S (1989) The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea*. Nature. 341: 164-166.

Kim, S.H., Lee, K.H., Kim, D.Y., Kwak, E., Kim, S and Kim, K.T (2015) Rhythmic control of mRNA stability modulates circadian amplitude of mouse *Period3* mRNA. J Neurochem. 132: 642-656.

Kim, T.D., Woo, K.C., Cho, S., Ha, D.C., Jang, S.K and Kim, K.T (2007) Rhythmic control of AANAT translation by hnRNP Q in circadian melatonin production. Genes Dev. 21: 797-810.

Kitayama, Y., Nishiwaki, T., Terauchi, K and Kondo, T (2008) Dual KaiC-based oscillations constitute the circadian system of cyanobacteria. Genes Dev. 22: 1513-1521.

Koike, N., Yoo, S.H., Huang, H.C., Kumar, V., Lee, C., Kim, T.K and Takahashi, J.S (2012) Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. Science. 338: 349-354.

Kojima, S., Gatfield, D., Esau, C.C and Green, C.B (2010) MicroRNA-122 modulates the rhythmic expression profile of the circadian deadenylase Nocturnin in mouse liver. PLoS One. 5: e11264.

Kojima, S and Green, C.B (2015) Circadian genomics reveal a role for post-transcriptional regulation in mammals. Biochemistry. 54: 124-133.

Konopka and Benzer (1971) Clock Mutants of *Drosophila melanogaster*. Proc Natl Acad Sci. 68: 2112-2116.

Kozak, M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15: 8125-8148.

Kozak, M (2002) Pushing the limits of the scanning mechanism for initiation of translation. Gene. 299: 1-34.

Krajnak, K., Dickenson, L and Lee, T.M (1997) The induction of Fos-like proteins in the suprachiasmatic nuclei and intergeniculate leaflet by light pulses in degus (*Octodon degus*) and rats. J Biol Rhythms. 12: 401-412.

Kronfeld-Schor, N and Einat, H (2012) Circadian rhythms and depression: human psychopathology and animal models. Neuropharmacology. 62:101-114.

Kwon, Y.J., Park, M.J., Kim, S.G., Baldwin, I.T and Park, C.M (2014) Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in Arabidopsis. BMC Plant Biol. 14:136.

Lee, H.J., Kim, L., Kang, S.G., Yoon, H.K., Choi, J.E., Park, Y.M., Kim, S.J and Kripke, D.F (2011) PER2 variation is associated with diurnal preference in a Korean young population. Behav Genet. 41: 273-277.

Lee, J.H., Kim, S.J., Lee, S.Y., Hwang, J-W and Suh, I.B (2015) Association of PER2 and CRY1 Polymorphisms with the Morningness-Eveningness in Korean Adults. Sleep Med Res. 6: 60-65.

Lee, K.H., Woo, K.C., Kim, D.Y., Kim, T.D., Shin, J., Park, S.M., Jang, S.K and Kim, K.T (2011) Rhythmic interaction between *Period1* mRNA and hnRNP Q leads to circadian time-dependent translation. Mol Cell Biol. 32: 717-728.

Lee, S., Liu, B., Lee, S., Huang, S.X., Shen, B and Qian, S.B (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci. 109: E2424-E2432.

Lee, S.J., Liu, T., Chattoraj, A., Zhang, S.L., Wang, L., Lee, T.M., Wang, M.M and Borjigin, J (2009) Posttranscriptional regulation of pineal melatonin synthesis in *Octodon degus*. J Pineal Res. 47: 75-81.

Lee, T.M (2004) *Octodon degus*: a diurnal, social, and long-lived rodent. ILAR J. 45: 14-24.

Leloup, J.C (2009) Circadian clocks and phosphorylation: Insights from computational modelling. Cent Eur J Biol. 4: 290-303.

Leonard, S (1995) *In vitro* transcription of cRNA from PCR generated DNA fragments. Methods in Neuroscience. 26: 243-251.

Li, S., Wang, M., Ao, X., Chang, A.K., Yang, C., Zhao, F., Bi, H., Liu, Y., Xiao, L and Wu, H (2013) CLOCK is a substrate of SUMO and sumoylation of CLOCK upregulates the transcriptional activity of estrogen receptor-α. Oncogene. 32: 4883-4891.

Lipton, J.O., Yuan, E.D., Boyle, L.M., Ebrahimi-Fakhari, D., Kwiatkowski, E., Nathan, A., Güttler, T., Davis, F., Asara, J.M and Sahin, M (2015) The Circadian Protein BMAL1 Regulates Translation in Response to S6K1-Mediated Phosphorylation. Cell. 161: 1138-1151.

Lobanov, A.V., Turanov, A.A., Hatfield, D.L and Gladyshev, V.N (2010) Dual functions of codons in the genetic code. Crit Rev Biochem Mol Biol. 45: 257-265.

Mammen, A.P and Jagota , A (2011) Immunocytochemical evidence for different patterns in daily rhythms of VIP and AVP peptides in the suprachiasmatic nucleus of diurnal *Funambulus palmarum*. Brain Res. 1373: 39-47.

Marcheva, B., Ramsey, K.M., Peek, C.B., Affinati, A., Maury, E and Bass, J (2013) Circadian clocks and metabolism. Handb Exp Pharmacol. 217:127-155.

Masri, S., Zocchi, L., Katada, S., Mora, E and Sassone-Corsi, P (2012) The circadian clock transcriptional complex: metabolic feedback intersects with epigenetic control. Ann N Y Acad Sci. 1264: 103-109.

Mattam, U and Jagota, A (2014) Differential role of melatonin in restoration of age-induced alterations in daily rhythms of expression of various clock genes in suprachiasmatic nucleus of male Wistar rats. Biogerontology. 15: 257-268.

Mattam, U and Jagota, A (2015) Daily rhythms of serotonin metabolism and the expression of clock genes in suprachiasmatic nucleus of rotenone-induced Parkinson's disease male Wistar rat model and effect of melatonin administration. Biogerontology. 16: 109-123.

Matthaei, J.H., Voigt, H.P., Heller, G., Neth, R., Schoch, G., Kubler, H., Amelunxen, F., Sander, G and Parmeggiani, A (1966) Specific interactions of ribosomes in decoding. Cold Spring Harb Symp Quant Biol. 31: 25-38.

Maury, E., Ramsey, K.M and Bass, J (2010) Circadian rhythms and metabolic syndrome: from experimental genetics to human disease. Circ Res. 106: 447-462.

Mauvoisin, D., Wang, J., Jouffe, C., Martin, E., Atger, F., Waridel, P., Quadroni, M., Gachon, F and Naef, F (2014) Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. Proc Natl Acad Sci. 111: 167-172.

McGlincy, N.J., Valomon, A., Chesham, J.E., Maywood, E.S., Hastings, M.H and Ule, J (2012) Regulation of alternative splicing by the circadian clock and food related cues. Genome Biol. 13: R54.

Mehra, A., Baker, C.L., Loros, J.J and Dunlap, J.C (2009) Post-translational modifications in circadian rhythms. Trends Biochem Sci. 34: 483-490.

Meijer, J.H., Rusak, B and Harrington, M.E (1989) Photically responsive neurons in the hypothalamus of a diurnal ground squirrel. Brain Res. 501: 315-323.

Metzenberg, S (2007) Working with DNA. Taylor & Francis Group: Florence, KY, USA.

Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H and Miura, K (2008) Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. Nucleic Acids Res. 36: 861-871.

Nakahata, Y., Yoshida, M., Takano, A., Soma, H., Yamamoto, T., Yasuda, A., Nakatsu, T and Takumi, T (2007) A direct repeat of E-box-like elements is required for cell-autonomous circadian rhythm of clock genes. BMC Mol Biol. 9: 1.

Nakajima, M., Imai, K., Ito, H., Nishiwaki, T., Murayama, Y., Iwasaki, H., Oyama, T and Kondo, T (2005) Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation *in vitro*. Science. 308: 414-415.

Ohsaki, K., Oishi, K., Kozono, Y., Nakayama, K., Nakayama, K.I and Ishida, N (2008) The role of {beta}-TrCP1 and {beta}-TrCP2 in circadian rhythm generation by mediating degradation of clock protein PER2. J Biochem. 144: 609-618.

Okamura, H., Kawakami, F., Tamada, Y., Geffard, M., Nishiwaki, T., Ibata, Y and Inouye, S.T (1995) Circadian change of VIP mRNA in the rat suprachiasmatic nucleus following p-chlorophenylalanine (PCPA) treatment in constant darkness. Brain Res Mol Brain Res. 29: 358-364.

O'Neill, J.S and Reddy, A.B (2012) The essential role of cAMP/Ca2+ signalling in mammalian circadian timekeeping. Biochem Soc Trans. 40: 44-50.

Owens, N.C., Hess, F.M and Badoer, E (2003) In situ hybridization using riboprobes on free-floating brain sections. Methods Mol Biol. 326: 163-171.

Park, S.G., Hannenhalli, S and Choi, S.S (2014) Conservation in first introns is positively associated with the number of exons within genes and the presence of regulatory epigenetic signals. BMC Genomics. 15: 526.

Peabody, D.S (1987) Translation initiation at an ACG triplet in mammalian cells. J Biol Chem. 262: 11847-11851.

Pisarev, A.V., Kolupaeva, V.G., Pisareva, V.P., Merrick, W.C., Hellen, C.U and Pestova, T.V (2006) Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. Genes Dev. 20: 624-636.

Preußner, M and Heyd, F (2016) Post-transcriptional control of the mammalian circadian clock: implications for health and disease. Pflugers Arch. 468: 983-991.

Reddy, M.Y and Jagota A (2015) Melatonin has differential effects on age-induced stoichiometric changes in daily chronomics of serotonin metabolism in SCN of male Wistar rats. Biogerontology. 16: 285-302.

Reddy, V.D and Jagota, A (2014) Effect of restricted feeding on nocturnality and daily leptin rhythms in OVLT in aged male Wistar rats. Biogerontology. 15: 245-256.

Redlin, U and Mrosovsky, N (2004) Nocturnal activity in a diurnal rodent (*Arvicanthis niloticus*): the importance of masking. J Biol Rhythms. 19: 58-67.

Regier, J.C., Fang, Q.Q., Mitter, C., Peigler, R.S., Friedlander, T.P and Solis, M.A (1998) Evolution and phylogenetic utility of the period gene in Lepidoptera. Mol Biol Evol. 15: 1172-1182.

Reischl, S and Kramer, A (2011) Kinases and phosphatases in the mammalian circadian clock. FEBS Letters. 585: 1393-1399.

Rey, G and Reddy, A.B (2013) Connecting cellular metabolism to circadian clocks. Trends Cell Biol. 23: 234-241.

Ripperger, J.A and Albrecht, U (2012) The circadian clock component PERIOD2: from molecular to cerebral functions. Prog Brain Res. 199: 233-245.

Ripperger, J.A and Brown, S.A (2010) Transcriptional regulation of circadian clocks. In: Albrecht, U. (Ed.), The circadian clock. Springer, New York, Dordrecht, Heidelberg, London. pp. 37-78.

Ripperger, J.A and Merrow, M (2011) Perfect timing: epigenetic regulation of the circadian clock. FEBS Lett. 585: 1406-1411.

Rissman, E.F (2004) Thinking Outside the Mouse Box: The Importance of Comparative Laboratory Animal Models in Research. ILAR J. 45: 1-3.

Roll.U., Dayan,T and Kronfeld-Schor, N (2006). On the role of phylogeny in determining activity patterns of rodents. Evol Ecol. 20: 479-490.

Sambrook, J and Russel, D.W (2001) Molecular cloning: A laboratory manual. 3rd ed, Cold Spring Harbour Laboratory Press, New York.

Schumann, D.M., Cooper, H.M., Hofmeyr, M.D and Bennett, N.C (2006) Light-induced *Fos* expression in the suprachiasmatic nucleus of the four-striped field mouse, *Rhabdomys pumilio*: A southern African diurnal rodent. Brain Res Bull. 70: 270-277.

Shigeyoshi, Y., Taguchi, K., Yamamoto, S., Takekida, S., Yan, L., Tei, H., Moriya, T., Shibata, S., Loros, J.J., Dunlap, J.C and Okamura, H (1997) Light-induced resetting of a mammalian circadian clock is associated with rapid induction of the *mPer1* transcript. Cell. 91: 1043-1053.

Silver, R and Rainbow, M (2013) The Suprachiasmatic Nucleus and the Circadian Timekeeping System of the Body. D.W. Pfaff (Ed.), Neuroscience in the 21st Century, Springer, NY. pp. 1847-1888.

Smale, L., Lee, T and Nunez, A.A (2003) Mammalian diurnality: some facts and gaps. J Biol Rhythms. 18: 356-366.

Spanagel, R., Rosenwasser, A.M., Schumann, G and Sarkar, D.K (2005) Alcohol consumption and the body's biological clock. Alcohol Clin Exp Res. 29: 1550-1557.

Stojkovic, K., Wing, S.S and Cermakian, N (2014) A central role for ubiquitination within a circadian clock protein modification code. Front Mol Neurosci. 7: 69.

Takahashi, J.S., Hong, H.K., Ko, C.H and McDearmon, E.L (2008) The genetics of mammalian circadian order and disorder: implications for physiology and disease. Nat Rev Genet. 9: 764-775.

Takahashi, K., Maruyama, M., Tokuzawa, Y., Murakami, M., Oda, Y., Yoshikane, N., Makabe, K.W., Ichisaka, T and Yamanaka, S (2005) Evolutionarily conserved non-AUG translation initiation in NAT1/p97/DAP5 (EIF4G2). Genomics. 85: 360-371.

Tarragon, E., Lopez, D., Estrada, C., Ana, G.C., Schenker, E., Pifferi,F., Bordet, R., Richardson, J.C and Herrorro, M.T (2013) *Octodon degus*: A model for the cognitive impairment associated with Alzheimer's Disease. CNS Neurosci Ther. 19: 643-648.

Tauber, E., Roe, H., Costa, R., Hennessy, J.M., Kyriacou, C.P (2003) Temporal Mating isolation driven by a behavioural gene in Drosophila. Curr Biol. 13: 140-145.

Taylor, B.L and Zhulin, I.B (1999) PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev. 63: 479-506.

Tikole, S and Sankarasubramanian, R (2012) A survey of mRNA sequences with anon-AUG start codon in RefSeq database. J Biomol Struct Dyn. 24: 33-42.

Touriol, C., Bornes, S., Bonnal, S., Audigier, S., Prats, H., Prats, A.C and Vagner, S (2003) Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. Biol Cell. 95: 169-178.

Ueda, H.R., Hayashi, S., Chen, W., Sano, M., Machida, M., Shigeyoshi, Y., Iino, M and Hashimoto, S (2005) System-level identification of transcriptional circuits underlying mammalian circadian clocks. Nat Genet. 37: 1872-1892.

Ukai-Tadenuma, M., Yamada, R.G., Xu, H., Ripperger, J.A., Liu, A.C and Ueda, H.R (2010) Delay in feedback repression by Cryptochrome 1 is required for circadian clock function. Cell. 144: 268-281.

Urrutia,R., McNiven, M.A and Kachar, B (1993) Synthesis of RNA probes by the direct *in vitro* transcription of PCR-generated DNA templates. J Biochem Biophys Methods. 26: 113-120.

VanderLeest, H.T., Houben, T., Michel, S., Deboer, T., Albus, H., Vansteensel, M.J., Block, G.D and Meijer, J.H (2007) Seasonal encoding by the circadian pacemaker of the SCN. Curr Biol. 17: 468-473.

Vanselow, K., Vanselow, J.T., Westermark, P.O., Reischl, S., Maier, B., Korte, T., Herrmann, A., Herzel, H., Schlosser, A and Kramer, A (2006) Differential effects of PER2 phosphorylation: molecular basis for the human familial advanced sleep phase syndrome (FASPS). Genes Dev. 20: 2660-2672.

Vivanco, P., Otalora, B.B., Rol, M.A and Madrid, J.A (2010) Dissociation of the circadian system of *Octodon degus* by T28 and T21 light-dark cycles. Chronobiol Int. 27: 1580-1595.

Vivanco, P., Rol, M.A and Madrid, J.A (2010) Pacemaker phase control versus masking by light: setting the circadian chronotype in dual *Octodon degus*. Chronobiol Int. 27: 1365-1379.

Wegrzyn, J.L., Drudge, T.M., Valafar, F and Hook, V (2008) Bioinformatic analyses of mammalian 5'-UTR sequence properties of mRNAs predicts alternative translation initiation sites. BMC Bioinformatics. 9: 232.

Weinreb, D.B (2001) Proposed Strategies for Using Non-Coding Sequences, Pseudogenes and Interspersed Elements to Construct Molecular Phylogenies. http://bioinfo.mbb.yale.edu

Wulff, K., Gatti, S., Wettstein, J.G and Foster RG (2010) Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. Nat Rev Neurosci. 11: 589-599.

Xu, Q., Walker, D., Bernardo, A., Brodbeck, J., Balestra, M.E and Huang, Y (2008) Intron-3 retention/splicing controls neuronal expression of apolipoprotein E in the CNS. J Neurosci. 28: 1452-1459.

Xu, Y., Toh, K.L., Jones, C.R., Shin, J.Y., Fu, Y.H and Ptácek, L.J (2007) Modeling of a human circadian mutation yields insights into clock regulation by PER2. Cell. 128: 59-70.

Yagita, K., Tamanini, F., Yasuda, M., Hoeijmakers, J.H., van der Horst, G.T and Okamura, H (2002) Nucleocytoplasmic shuttling and mCRY-dependent inhibition of ubiquitylation of the mPER2 clock protein. EMBO J. 21: 1301-1314.

Yan, L and Silver, R (2002) Differential induction and localization of *mPer1* and m*Per2* during advancing and delaying phase shifts. Eur J Neurosci. 16: 1531-1540.

Yang, Y., Duguay, D., Fahrenkrug, J., Cermakian, N and Wing, S.S (2014) USP2 regulates the intracellular localization of PER1 and circadian gene expression. J Biol Rhythms. 29: 243-256.

Yoo, S.H., Mohawk, J.A., Siepka, S.M., Shan, Y., Huh, S.K., Hong, H.K., Kornblum, I., Kumar, V., Koike, N., Xu, M., Nussbaum, J., Liu, X., Chen, Z., Chen, Z.J., Green, C.B and Takahashi, J.S (2013) Competing E3 ubiquitin ligases govern circadian periodicity by degradation of CRY in nucleus and cytoplasm. Cell. 152: 1091-1105.

Yue, Y., Liu, J and He, C (2015) RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. Genes Dev. 29: 1343-1355.

Zee, P.C., Attarian, H and Videnovic, A (2013) Circadian rhythm abnormalities. Continuum (Minneap Minn). 1: 132-147.

Zhang, L., Weng, W and Guo, J (2011) Posttranscriptional mechanisms in controlling eukaryotic circadian rhythms. FEBS letters. 585: 1400-1405.

Zhou, M., Kim, J.K., Eng, G.W.L., Forger, D.B and Virshup, D.M (2015) A Period2 Phosphoswitch Regulates and Temperature Compensates Circadian Period. Mol Cell. 60: 77-88.

Zhu, H., Vadigepalli, R., Rafferty, R., Gonye, G.E., Weaver, D.R and Schwaber, J.S (2012) Integrative gene regulatory network analysis reveals light-induced regional gene expression phase shift programs in the mouse suprachiasmatic nucleus. PLoS One. 7: e37833.

APPENDIX

AI. LIST OF CHROMATOGRAMS, BLAST, CLUSTAL ALIGNMENT, CONSERVED DOMAIN SEARCH REPORTS AND NUCLEOTIDE SEQUENCE DOCUMENTS ATTACHED

	File Name
C1	CON_E R113_D01_2014.02.20
C2	P2_INT1_M13F_E06_2014.11.05
D1	Nucleotide sequences exported from C1 and C2
A1	Alignment of 3 sequencing data (a,b,c) of F.palmarum gDNA FconE/R113 PCR gel eluates with TA clone with -140P2 intron 1F- R113 fragment to generate consensus sequence
B1	NCBI-BLAST report of consensus sequence- hits to Per2 is enlisted
C3	DNA-M_R113_F02_2014.05.16
C4	DNA_WM_R113_G11_2014.06.16
D2	Nucleotide sequence exported from C3,C4
B2	NCBI-BLAST report of C3 - specific hits to mPer2 is enlisted
B3	NCBI-BLAST report of C4 - specific hits to mPer2 is enlisted
A2	Alignment of C57BL/6 and wild mouse sequence with GenBank sequences for C57BL/6 and *R.norvegicus*
A3	Alignment of sequences flanking the Per2 'start' codon in mammals with that of *F.palmarum*
C5	SquirrelDNA_F84 D01.ab1.
C6	SquirrelDNA Sq.1_F_E03.ab1.
C7	P2TA-2_M13F_F04_2014.12.09
C8	P2TA-2_F581_G01_2015.06.25
C9	P2TA-2_R610_H01_2015.06.25
C10	P2TA-2_F927_B04_2015.06.20
C11	P2TA-2_F2542_H11_2015.06.11
C12	P2TA-2_R2542_A12_2015.06.12
C13	P2TA-2_R2424_B12_2015.06.12
C14	P2TA-2_F1517_B02_2015.06.25
C15	P2TA-2_R2216_E04_2015.06.20
C16	p23a_F2626_G06_2016.02.16
C17	P2TA-2_R2784_C12_2015.06.12
C18	P2TA-2_R2972_D04_2015.06.20
C19	p23a_F3040_H06_2016.02.16
C20	P2TA2_R3316_C04_2015.06.20
C21	p23a_F3386_A07_2016.02.16.
C22	P2TA2_MINN3_A02_2015.06.25 (primer minus3 R3'UTR)
C23	P2TA2_1R3primerUTR_E01_2015.05.21 (primer minus1 R3'UTR)
C24	P2TA-2_M13R_G04_2014.12.09
D3	Complete nucleotide sequence of *F.palmarum Per2* derived from C7 to C 24 Expasy Translate Report. NCBI Conserved Domain Search Report.
A4.	CLUSTAL alignment of *F.palmarum* aminoacid sequence with that of *Mus musculus, Rattus norwegicus, Mesocricetus auratus, Cricetulus griseus, Ictidomys tridecemlineatus, Homo sapiens.*The FASP domain is highlighted

C25	P2TA_F927_D04_2014.12.09
C26	P2TA_R2784_A01_2015.05.21 (reverse transcribed)
A5	CLUSTAL alignment of *F.palmarum Per2* and *Per2v.*
D4	Predicting the region of possible Intronic retention and Multiple exon deletion in *F.pal per2* Expasy Translate tool report. NCBI Conserved domain Search report.
C27	Vip_M13R_C07_2016.02.16
D5	Nucleotide sequence exported from C27 NCBI BLAST report.
C28	*Per1*
C29	*Per3*
C30	*Cry1*
C31	*Cry2*
C32	*Bmal1*
C33	Ham ConE +R1
C34	Gui ConE +R1
C35	G-gdna_Fcone_H07_2016.02.16
C36	HgDNA_Fcone_D07_2016.02.16
C37	DNAFP_ConE_D11_2014.06.16
A6	Alignment of 4 sequencing reactions of *F.palmarum* EE box and 1st intronic sequence

A II. LIST OF ABBREVIATIONS

AANAT	Aralkylamine N-acetyltransferase
Aβ-APP	Amyloid precursor protein
AHA	Anterior Hypothalamic Area
AMPK	5' adenosine monophosphate-activated protein kinase
ARNT	Arylhydrocarbon receptor nuclear translocator
Arntl	Aryl Hydrocarbon Receptor Nuclear Translocator Like
ATP	Adenosine triphosphate
aTIS	alternative Translation Initiation Site
bHLH	Basic helix-loop-helix
Bmal1	Brain and muscle Arnt-like protein-1
BSA	Bovine Serum Albumin
BST	Bed Nucleus Of Stria Terminalis
βTRC	beta-transducin repeat containing E3 ubiquitin protein ligase
bZIP	The Basic Leucine Zipper Domain
CaCl2	Calcium Chloride
CALB	Calbindin
CAMK	Ca2+/calmodulin-dependent protein kinase
CCA1	Circadian Clock Associated 1
CCG	Clock Control gene
cDNA	Complementary Deoxyribose Nucleic Acid
cGMP	Cyclic guanosine monophosphate
CICR	Calcium Induced Calcium Release
CK	Casein Kinase
CLD	Cytoplasmic Localization Domain
CLK	Clock
Clock	Circadian locomoter output cycles protein kaput
cmyc	V-Myc Avian Myelocytomatosis Viral Oncogene Homolog
COS1	CV-1 (simian) in Origin, and carrying the SV40 genetic material
CRE	cAMP response element
CREB	cAMP-response element binding protein
CRY	Cryptochrome
CT	Circadian Time
Cul1	Cullin 1
CWO	clockwork orange
CYC	Cycle
DBP	D site of albumin promoter (albumin D-box) binding protein
DBPE	DBP-binding element
DD	Constant dark
DEC 1,2	Deleted In Esophageal Cancer 1
DEPC	Diethylpyrocarbonate
DMH	Dorsomedial Hypothalamus
dNTP	Deoxyribonucleotide triphosphates
DTT	Dithiothreitol

DYRK 1A	Dual Specificity Tyrosine Phosphorylation Regulated Kinase 1A
4E-BP1,4	Eukaryotic translation initiation factor 4E-binding protein 1,4
EDTA	Ethylenediaminetetraacetic acid
ELF 3, 4	Early Flowering 3,4
eIF	Eukaryotic Initiation Factor
ERα	Estrogen Receptorα
EZH2	Enhancer of zeste homolog 2
FASPS	Familial advanced sleep-phase syndrome
Fbxl3	F-Box And Leucine-Rich Repeat Protein 3
FRH	FRQ-interacting RNA helicase
FRQ	Frequency
FTO	Fat Mass and Obesity Associated Protein
GABAergic	Gamma-Aminobutyric Acid
Gapdh	Glyceraldehyde 3-phosphate dehydrogenase
GC	Guanylate cyclase
gDNA	Genomic DNA
GHT	Geniculo Hypothalamic Tract
GRP	Gastrin-releasing peptide
GSK	Glycogen Synthase Kinase
H3	Histone 3
HAT	Histone Acetyl Transferase
HDAC	Histone Deacetylase
hnRNP	Heterogeneous nuclear ribonucleoproteins
Hnf4α	Hepatocyte nuclear factor 4 alpha
IGL	Intergeniculate leaflet
IR	Intron Retention
IRES	Internal Ribosome Entry Sites
ISH	In situ Hybridization
JmJ D-domain	Jumonji D-domain
KDM 5A	Lysine demethylases 5A
KoAc	Potassium Acetate
LB	Lysogeny Broth
LHY	Late Elongated Hypocotyl
LncRNA	Long noncoding RNA
LSD1	lysine (K)-specific demethylase 1A
MAPK	Mitogen-activated protein kinases
M & E	Morning and Evening

METTL3	Methyltransferase Like 3
miRNA	Micro RNA
MLL	Mixed Lineage Leukemia
MnCl2	Manganese Chloride
MPOA	Medial Preoptic Area
MYBBPIA	Myb-binding protein 1a
NaCl	Sodium Chloride
NAD+	Nicotinamide adenine dinucleotide
NAMPT	Nicotinamide phosphoribosyltransferase
NCBI	National Center for Biotechnology Information
NES	Nuclear Export Signal
NGS	Next-Generation Sequencing
NLK	Nemo Like Kinase
NLS	Nuclear Localization Signal
NMD	Nonsense Mediated Decay
NMO	Nemo like kinase
NO	Nitric Oxide
Noc	Nocturin
NOS	Nitric oxide synthase
Npas2	Neuronal PAS Domain Protein 2
NPY	Neuropeptide Y
Nurr1	The Nuclear receptor related 1 protein
OC	Optic Chiasm
OD	Optical Density
PACAP	Pituitary adenylate cyclase-activating peptide
PAG	Periaqueductal grey
PARP	Poly ADP ribose polymerase
PAR	
PCR	Polymerase Chain Reaction
PDP1	Pyruvate dehyrogenase phosphatase catalytic subunit 1
PER 1,2,3	Period 1,2,3
PHA	Posterior hypothalamic area
PHI	peptide histidine-isoleucin
PKA	Protein Kinase A
PKC	Protein kinase C
PKG	Protein Kinase C
PML	Promyelocytic Leukemia
POA	Preoptic Area
Poly (A)	Poly adenosine
PP 1,2A,4,5	Protein Phosphatase
PPARα	Peroxisome proliferator-activated receptor alpha
PRR 3,5,7,9	
PSF	Splicing factor, proline- and glutamine-rich

PT	paratanial nucleus
PTB	Polypyrimidine tract Binding protein
PTC	Premature Termination Codon
PTM	Posttranslational modification
PTO	Posttranslational oscillator
PVH	Paraventricular nucleus of hypothalamus;
PVT	Paraventricular nucleus
RACE	Rapid Amplification of cDNA ends
RBP	RNA Binding Protein
Rch	retrochiasmatic area;
Rev-Erb α, β	Reverse-Erythroblastosis α, β
RHT	Retinohypothalamic tract
RNA	Ribonucleic Acid
RNA Pol II	RNA Polymerase II
RNMT	RNA (Guanine-7-) Methyltransferase
ROR α, β	Retinoid-related orphan receptors α, β
RRE	Rev response element
RT	Room Temperature
RUST	Regulated Unproductive Splicing and Translation
SCN	Suprachiasmatic nuclei
SDS	Sodium Dodecyl Sulfate
SGG/GSK	SHAGGY/Glycogen Synthase Kinase
SIM	Single Minded Protein
SIRT1	silent mating type information regulation 2 homolog 1
Skp1-Cul1-Fbxl3	S-Phase Kinase-Associated Protein 1-Cullin 1-Fbxl3
SLIMB	Supernumerary limbs
SNP	Single Nucleotide polymorphism
SP	Substance P
SS	Somatostatin
SSC	Saline Sodium Citrate
SUMO	Small Ubiquitin-like Modifier
TE buffer	Tris EDTA
TEF1	Transcription Enhancer Factor 1
TF	Transcription Factor
TIM	Timeless
TOC1	Timing of CAB Expression 1
TORC1	Transcriptional coactivator for CREB1
TRα	Thyroid Hormone Receptor α
Tris-Cl	Tris- Chloride
TTLF	Transcriptional Translational Feed back
uORF	upstream Open Reading Frame

UTR	Untranslated Region
vLGN	Ventral Lateral geniculate nucleus
Vip	Vasoactive Intedtinal Peptide
VMH	Ventromedial nucleus of the hypothalamus
VRI	Vrille
WDR5 adapter	WD Repeat Domain 5
WC	White Collar
WD	tryptophan-aspartic acid
WTAP	WT1-Associated protein
WCC	White Collar Complex
ZT	Zeitgeber Time

A III. LIST OF FIGURES

CLOCK (C) and BMAL (B) dimerize and bind to E-boxes in the promoter region of Period (Per) genes, Cryptochrome (Cry) genes and Rev-Erb*a*, activating their transcription. SCN shell neurons communicate with SCN targets using VP and GABA as neurotransmitters. Additionally, the SCN communicates with some target sites using a diffusible signal. (b) Neuropeptides of the SCN. The SCN of the hamster can be divided into a ventrolateral core and a dorsomedial shell, both densely packed with somata of small neurons. The main neuropeptide transmitters of these areas are vasoactive intestinal polypeptide (VIP), peptide histidine-isoleucin (PHI), gastrin-releasing peptide (GRP), calbindin (CalB), somatostatin (SS), and vasopressin (VP). Most neurons contain additionally GABA. The somata of the neurons are shown as larger circles, their projections as small circles.

Fig.7. Diagrammatic representation of a photo inducible clock gene with its promoter elements, CRE, E-box and D element. Light activates signaling cascade downstream of Glutamate induction. Based on the temporal profile of the photic cue the signaling bifurcates at the level of Nitric Oxide (NO) production, both the phase advancing and delaying light pulse activate CRE mediated transcript induction.

Fig.8. Post-transcriptional mechanisms acting on the (pre) mRNA to regulate the mammalian circadian clock.

Fig.9. Schematic representation of the main components, kinases and phosphatases and their interaction in the molecular circadian clock of Neurospora, Drosophila and mammals. Transcriptional repressors in the molecular clockwork are depicted in light blue, transcriptional activators in green. Kinases and phosphatases are coloured according to their effect on circadian period. When inhibition of kinase/phosphatase activity causes period lengthening, shortening or arrhythmia, the kinase is depicted in yellow, orange or purple respectively. With no or unclear effect on period the kinases/phosphatases are depicted in grey. Lines between components represent interactions between kinase/phosphatase and its substrates. (A) Neurospora): PP1 – protein phosphatase 1, PP2A – protein phosphatase 2A, PP4 – protein phosphatase 4, PKA – protein kinase A, CK1-a – casein kinase Ia, CAMK-1 – calcium/calmodulin-dependent protein kinase, CKII – casein kinase 2, FRQ – frequency, WCC – white collar complex. (B) (Drosophila): PP1– protein phosphatase 1, PP2A – protein phosphatase 2A, DBT/CKI – DOUBLETIME/casein kinase I, CK2 – casein kinase 2, SGG/GSK-3 – SHAGGY/glycogen synthase kinase 3, PER – PERIOD, TIM – TIMELESS, CLK – CLOCK. (C) (mammals): PP1 –protein phosphatase 1, PP5 – protein phosphatase 5, AMPK – adenosine monophosphate- activated protein kinase, CKI – casein kinase I, CK2 – casein kinase 2, GSK-3 – glycogen synthase kinase 3, PERs – PERIOD proteins 1-3, CRYs – CRY proteins 1 and 2, BMAL1 – brain and muscle aryl hydrocarbon receptor nuclear translocator (ARNT)-like 1),CLOCK – circadian locomotor output cycles kaput, REVERB-a – nuclear receptor subfamily 1, group D, member 1.

Fig.10. Model for the differential effects of PER2 phosphorylation on circadian oscillations. PER2 contains at least two functionally different sets of phosphorylation sites—one primarily mediating proteasomal degradation (green), the other nuclear retention (purple). In FASPS-PER2 (*right* side of the panels), the latter cannot be

phosphorylated because Ser 659 is mutated to glycine. The region responsible for nuclear retention cannot be phosphorylated (red crosses), leading to premature nuclear export of the PER2–CRY complex, and thus to an earlier cytosolic degradation and to a faster circadian cycle.

Fig.11. Model of dynamic chromatin transitions at circadian genes over the 24 h day (day, open bar at bottom; night, closed bar at bottom). Activation of BMAL1 and CLOCK regulated genes coincides with acetylation of histone residues by HATs, and tri-methylation of lysine 4 of histone H3 by Set2/Ash1 (via a WDR5 adapter) or MLL, and the removal of repressive histone methylations by LSD1 or JmjD-domain proteins. These activities are counterbalanced by the recruitment of HDACs or SIRT1, and LSD1 and JmjDdomain proteins, respectively. Similar scenarios may occur at RORa/REV-ERBa and DBP regulated circadian target genes as well. Indeed, REV-ERBa was recently identify as the main anchor point for HDAC3.

Fig.12. A minimal circuit model for the mammalian circadian transcriptional network. Morning (E/E') box and (D-box) and night (RRE) DNA elements lie at the heart of the circadian clock Solid lines indicate activating (green) and inhibitory (pink) interactions between clock gene types. The dotted lines indicate how *Cry 1* represses morning expression through the combined action of day and night sequences.

Fig.13. Regulation of PER2 intracellular dynamics plays a crucial role in core clock function. The core clock (inner circle, yellow) consists of a negative feedback loop driven by the transcription, translation, repression, and degradation of core clock components PER2 and CRY, which together drive the circadian rhythms of a cell. The outer circle in orange illustrates molecular mechanisms pertaining to PER2 intracellular dynamics, which in turn affect core clock events. Together, these processes take ~24 hours to complete. Abbreviations: Ac-acetyl group, B-BMAL1, C-CLOCK, E-E-box promoter elements, ub-ubiquitin. HDAC1- histone deacetylase 1. KDM5A-lysine (K)-specific demethylase 5A. PML-promyelocytic leukaemia protein. NLS-nuclear localization sequences, BTRC-beta-transducin repeat containing E3 ubiquitin protein ligase, NONO non-POU domain containing, octamer-binding, SIN3A-SIN3 transcription regulator homolog A, SIRT1-Sirtuin 1, CLD-cytoplasmic localization domain, NES-nuclear export sequence, PSF-Polypyrimidine tract binding protein-associated splicing factor, WDR5-WD repeat domain 5, MYBBP1A-MYB binding protein.

Fig.14. a) Transmission of circadian clock information to nuclear receptor-target genes by PER2. PER2 mediates primary output from the molecular oscillator. By direct interaction with the nuclear receptor (NR) homo- or heterodimers, it can affect the corresponding target gene promoters and metabolic or physiological processes. The activity of PER2 may be modulated by the input (e.g., light or food) to the circadian oscillator. (b) A Proposed Model for the Role of *mPer2* in tumour suppression. The solid lines indicate the pathways that have been demonstrated. The dashed line indicates a regulatory pathway(s) that is still not fully understood.

Fig.15. Representative actograms showing locomotory rhythms of a (a) nocturnal (b) diurnal and (c) crepuscular organism.

Fig.16. (a) *Funambulu palmarum* (South Indian three stripped palm squirrel) housed in a cage (b) Actogram showing the locomotory activity confined to the light phase of an LD cycle and to the subjective day under DD (c) The circadian neuro anatomy of the model (left panel) in comparison to nocturnal rat (d) VIP-ir peaks at ZT 6, at the SCN core region (left) whereas AVP-ir peaks at ZT 12 delineating the shell region of SCN.

Fig.17. (a) A diagrammatic representation pf Per2 mRNA indicating the sense and antisense primer region. (b) Agarose gel images showing PCR products (1) F84/R1146 (2) F927/R2216 (3) & (4) respective 2nd round PCR products.

Fig.18. Exonic architecture of Per2 constructed for *Homo sapiens, Rattus norvegicus* and *Mus musculus.*

Fig.19. Diagrammatic representation showing the E'E regulatory element and the antisense primer region.

Fig.20. (a) gDNA extracted from the SCN of *F.palmarum*. (b)10 kb amplicon obtained with the primers F cone/R113 in Long Range PCR with gDNA as template.(c) Amplicons after Gel extraction loaded on agarose gel

Fig.21. (a) Chromatogram with the 'CAC' read highlighted (b) image of BLAST report showing similarity to *H.sapien Per2* (c) image of aligned sequences from repeated experiments, CAC and GTG (when reverse transcribed) is highlighted.

Fig.22. amplicons from F cone/R113 PCR using c57bl6 (a) and wild mouse (b) SCN-gDNA template *F.palmarum* gDNA was used as positive control with both. (c) & (d) the gel extracted PCR products, from (a) and (b).

Fig.23. A segment of the chromatogram with the CAT (reverse transcribed as ATG) read highlighted in c56bl7 (a) and wild mouse (b). Short region of CLUSTAL alignment (c) with the lab generated sequencing data of c56bl7 (bl6 r113) and wild mouse (wildmusr113) compared with the *F. palmarum* sequence (fpr113,a,b,c) and NCBI data for the same region in mouse, rat and humans (ncbibl7, ncbirat, ncbihomo respectively).

Fig.24. (a) Diagrammatic representation of *F. palmarum Per 2* gene showing the primer region (b) colony PCR with gene specific primers for selection of TA plasmid colonies with the insert. (c) small region of chromatogram (plasmid sequencing with vector specific primers) highlighting the 'GTG' at the start codon position (d) alignment showing comparison of the region in *F.palmarum* [fp113 a,b and c, fpta (TA plasmid with the insert)] with *Homo sapiens* (hper2), *Macaca mullata* (macper2), *Bos taurus* (bosper2), *Arvicanthis* (arvper2), *Mus musculus* (musper2), *Rattus* (rper2). The intron-exon junction is 'boxed' in red.

Fig.25. (a) pictorial representation of *Per2* mRNA with the primer region indicated. (b) Gel with the PCR products primed using F Start 1(d)/R3'UTR944 (c) Colony PCR for positive selection of TA-XL plasmids carrying the insert. Each colony was analysed using 2 primer pairs characterized previously, F start1 (d)/R610 and F84/R1146.

Fig.26. Diagram indicating the approximate position of individual primers across the *Per2* transcript which was used for gene sequencing.

Fig.27. Diagrammatic representation of the *F.palmarum Per2* variant (*F.pal Per2v*)

Fig.28. Sense and antisense primer position depicted on pre-pro *Vip* polypeptide. (b) Agarose gel with the PCR product just below the 500bp range. (c) Colony PCR results from 6 different positive colonies after TA cloning.

Fig.29. Diagram showing the primers that are designed to distinguish between the variants of *Per2*

Fig.30. Agarose gel with specific amplicons that will be substrates for in vitro transcription

Fig.31. Integrity of cRNA probes confirmed by gel electrophoresis

Fig.32. PCR amplification of *F. palmarum* clock genes, agarose gel showing single specific amplification for each gene.

Fig.33. Histogram showing the expression of clock genes at ZT0 and ZT 12. X axis-Relative gene expression. Y axis-ZT time points.

Fig.34. *Per2* gene of *F. palmarum* depicting the primer region. (b) The PCR product of the primer combination F cone/Rintron1 from both SCN and cortex gDNA (C28)

Fig.35. (a) gDNA extracted from the SCN of Guinea pig and Hamster (b) PCR products with Hamster gDNA sample (c) PCR products with Guinea pig sample. (d) gel extracted amplicon from hamster.(e) gel extracted amplicon from hamster and guinea pig gDNA reaction. The size difference of the *Per2* intergenic sequence in these animals is reflected in the amplicon size.

A IV. LIST OF TABLES