Natural variation within a clock gene of *Drosophila melanogaster* : a phenotypic and molecular analysis

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Lesley Anne Sawyer
Department of Genetics, University of Leicester.

September 1996.

ACKNOWLEDGEMENTS

Firstly I would like to show my heartfelt appreciation to my parents for supporting and encouraging me, especially in this last year, without which 'it' would not have been possible.

I appreciate the time and enthusiasm that my supervisor Professor Bambos Kyriacou devoted to this project. Unfortunately being primarily a behavioural study he was unable to give me his hands on experience of ' change the buffer'. However, his helpful criticism and dreams of theories were always interesting to hear and inevitably led to more experiments!

I am indebted to many of my colleagues in lab 136 , to Alex Peixoto for supervision on my arrival in Leicester, to Clare Byrne for that dirty laugh and to Ezio Rosato for his technical assistance and a constant willingness to help. I would also like to thank Helen Parkinson and Roz Stanley, for keeping everything in perspective when things looked pear shaped, and Martin Couchman for his computer wizardry. I valued the cheerful support of the newer arrivals of the group, Ekly Ben-Shlomo and Anthony Rogers, and the old, Genevieve Herbert, whose e-mails from Oz gave great incentive to finish. In addition I thank all the rest of the members of lab 136, my undergraduate project student Anna Goostrey and the department for their help, advice and science.

Special thanks goes to members of the coughie club old and new, especially Ian, Wendy and Simon, for their friendship and their willingness to lend an ear at times of crisis.

I would also like to thank NERC for the studentship and Employment Services for the financial support in the final year.

Finally I'd like to thank all my friends from Leicester and afar, particularly Lettuce, Anna, Bridget, Sally and Vanessa.

ABBREVIATIONS

bp	base pair
CT	circadian time
DD	freerunning in constant darkness
h	hours
kb	kilobase pairs
LD 12:12	12 hours light/dark cylce
min	minutes
mRNA	messenger RNA
PCR	polymerase chain reaction
s	seconds
SEM	standard error of the mean
ZT	zeitgeiber time (ZT0 lights on)

ABSTRACT

The clock gene *period* encodes a repetitive series of Threonine-Glycine (Thr-Gly) pairs in *D. melanogaster*. This sequence is polymorphic in length and shows a clinal distribution in Europe, whereby the shorter length variants are more frequent in warmer regions and the longer ones in colder areas. A series of new correlations performed with the European population data suggests that thermal selection may be a factor in generating the cline. DNA sequence analysis of various Thr-Gly haplotypes supports a model where the $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ alleles are considered as ancestral, since all variants can be derived from them. A reciprocal cline is observed for the $(Thr\text{-}Gly)_{20}$ length variant in eastern Australia, suggesting that selection may be maintaining the polymorphism, and a further correlation of temperature with allele frequency suggests, as in European populations, that the cline may be thermally mediated. However, the other major allele the $(Thr\text{-}Gly)_{17}$ was spatially homogenous. The temperature compensation of the clock was examined by investigating the freerunning locomotor activity cycles of different natural Thr-Gly length variants from Europe and Australia. A relationship between Thr-Gly length and temperature compensation which could help explain the continental spatial differences was observed. Locomotor activity profiles in light/dark cycles were also examined in the European variants, and revealed subtle phenotypic differences between the Thr-Gly length variants that could also contribute to the observed clinal patterns. The energy expenditure of different Thr-Gly length variants was also measured and revealed behavioural adaptations for possible fitness advantages. Throughout the behavioural analyses there was also evidence that the structural $(Thr\text{-}Gly)_3$ motif played an important role in the temperature compensation mechanism and energy expenditure of the fly.

CONTENTS

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Biological rhythms

Within the cells or biological units of organisms, ranging from single celled bacteria to human beings, a multitude of clocks tick to maintain the daily cycles of biochemical, physiological and behavioural phenomena. The recurrence of any event within a biological system at more-or-less regular intervals can be considered a rhythm (Kalmus, 1935). The notion of a rhythm is sufficiently vague to describe a wide variety of phenomena, as diverse as observations of criminal repetition studied in reference to serial killers (Zagury, 1996; Hale, 1994). It has been suggested that the killer's actions sometimes represent a lethal cycle or 'tapeloop' (Ginsburg, 1993), playing out a series of acts and feelings, which come to an end, then begin again. However, unsupported by biogenic evidence this example can not be distinguished as a true biological rhythm. These can be either exogenous, reflecting a response to a periodic input from outside the biological unit, or endogenous, with the rhythm originating from within the biological unit by self sustaining oscillations. Several of these oscillations generate circadian cycles which persist with near 24 h periodicity under constant environmental conditions, their periods are stable over a physiological range of temperatures (Pittendrigh, 1974) and they can be reset by light or other environmental stimuli. A circadian pacemaking system consists of three basic components: a central clock or pacemaker, an input pathway, and the downstream outputs, which generate the biochemical and behavioural fluctuations.

Contemporary genetic studies of biological rhythms have involved the identification and isolation and application of mutations as well as the cloning and manipulation of DNA sequences. Some of the earliest circadian rhythms were found in eukaryotic microorganisms, for example *Chlamydomonas reinhardtii* (Bruce and Pittendrigh, 1957), whose rhythms range from phototactic responses (Bruce, 1970, 1972) to the mating competence of gametes (Demets *et al.*, 1987). One of the most intensively studied and best understood circadian systems is that of eclosion (adult emergence from the pupal case) of *D. pseudoobscura* and *D. melanogaster* (Pittendrigh, 1960, 1974). This model system was used to develop the formal properties of circadian oscillators, the mechanism of their entrainment to light cycles (Pittendrigh, 1965), and their role in photoperiodic measurement (Pittendrigh, 1981; Pittendrigh *et al.*, 1984). Some of the earlier studies of tractable single gene mutations that gave defective circadian phenotypes included Konopka and Benzer (1971) and Feldman and Hoyle (1973), who identified the *period* (*per*) locus in *Drosophila*

melanogaster and the *frequency* (*frq*) locus in *Neurospora crassa*, respectively. Subsequently, several groups have generated clock mutants in mice (*Clock*, Vitaterna *et al.*, 1994), serendipitously identified mutations in hamsters (*tau*, Ralph and Menaker, 1988), and used reporter genes to generate mutations in cyanobacteria (Kondo *et al.*, 1994) and plants (Millar *et al.*, 1995a).

Clock mutations other than *per* have been identified in *D. melanogaster*, for example *Toki* (Matsumoto *et al.*, 1994), *ritsu* (Murata *et al.*, 1995), *And* (Konopka *et al.*, 1991; Newby *et al.*, 1991), *psi-2*, *psi-3*, *gat* (Jackson, 1983), and *timeless* (Sehgal *et al.*, 1994) all affect the circadian rhythm. Furthermore, some mutations of the nervous system or of sensory organs have been reported to show abnormal circadian rhythms, for example *disco*, Dushay *et al.*, 1989, *glass*, Sehgal *et al.*, 1991 and *ebony*, Newby and Jackson, 1991.

Also the recent circadian and molecular characterisation of *Dreg-5* (*Drosophila* rhythmically expressed gene) shows mRNA expression in phase with *per* mRNA (as does *tim* mRNA, Sehgal *et al.*, 1995) under many different environmental conditions (Van Gelder and Krasnow, 1996). This latter novel circadianly expressed gene is dependant on *per* function. Analysis of *Dreg-5* and identification of additional *per*-dependent oscillating mRNA's provides a means of biochemically phenotyping the output pathways of the circadian pacemaker in *Drosophila*. Only the clock controlled genes (*ccg-1* and *ccg-2*) of *N. crassa* (Loros *et al.*, 1989, 1993) have been identified, where a genetic component of the circadian oscillator has also been characterised (*frq*) and thus can be placed in a genetic pathway for circadian rhythms. Both *ccg-1* and *ccg-2* appear to be partially dependent on *frq* for rhythmic expression (Arpaia *et al.*, 1993), but unfortunately they have not been characterised further in different environmental conditions.

1.2 Behavioural genetics of *per*

D. melanogaster has been employed extensively as a model genetic system for the analysis of circadian rhythms. The *per* gene was originally identified by screening for chemically induced mutations which altered the normal 24h circadian pupal-to-adult eclosion rhythm (Konopka and Benzer, 1971). The classic sex-linked mutations of the *per* gene abolish behavioural rhythmicity (per^{01}), lead to abnormally long, 28-30h (per^{Ll}) , or short, 19h (per^{s}), periods, perturb the clock resetting response (per^{s}), and disrupt temperature compensation (per^{Ll}) (Konopka *et al.*, 1989). The use of more powerful statistical techniques (Dowse and Ringo, 1989), has revealed significant single or multiple

periods in the ultradian domain (less than 24h), and thus per^{01} mutants are not considered strictly arrhythmic (Dowse *et al.*, 1987; Hamblen-Coyle *et al.*, 1989). These three mutations were later shown to affect a number of other phenotypes (reviewed in Hall, 1990; Hall and Kyriacou, 1990; Kyriacou, 1990; Dunlap, 1993) including the locomotor activity rhythm of adult flies (Konopka and Benzer, 1971), the 60 s ultradian cycle in the male's courtship song (Kyriacou and Hall, 1980; 1989; Kyriacou *et al.*, 1990b) and the 10 day infradian egg-to-adult developmental cycle (Kyriacou *et al.*, 1990a). Two additional short-period alleles, per^{T} and per^{Clk} generate behavioural rhythms with periods of 16-17h (Konopka *et al.*, 1995) and 22.5h (Dushay *et al.*, 1990), respectively.

per mutant heterozygotes, give semidominant effects (Konopka and Benzer, 1971; Smith and Konopka, 1981; 1982; Konopka *et al.*, 1995), while extra copies of per^{+} result in shorter periods, and fewer copies yields longer periods (Smith and Konopka 1982; Baylies *et al.*, 1987). These dosage effects suggested that per^{s} and per^{L1} are hypermorphic and hypomorphic alleles respectively (Smith and Konopka, 1982; Coté and Brody, 1986). Studies of the behavioural effects of targeted *per* mutations have indicated that shortened circadian periods can result from a range of *in vitro* single amino acid substitutions at or near the site of the original per^{s} mutation, or from insertion of a peptide antigen at the N-terminus (Rutila *et al.*, 1992; Baylies *et al.*, 1992).

One of the integral features of a biological clock is temperature compensation (Pittendrigh 1954; 1960), that is, the stability of period over a wide range of temperatures. In some *per* mutant strains, however, circadian period is temperature sensitive. The periods of per^{+} rhythms are constant between 15°C and 29°C; per^{s} periods are also constant, but slightly shorter, between 17°C and 29°C and very slightly longer at 15°C (Konopka *et al.*, 1989; Ewer *et al.*, 1990). Nevertheless, the periods of per^{L1} and per^{T} mutant rhythms are dramatically influenced by thermal challenges; per^{L1} rhythms become longer (Konopka *et al.*, 1989; Ewer *et al.*, 1990) and per^{T} rhythms become shorter (Konopka *et al.*, 1995) as the temperature increases from 15°C to 29°C. Thus, in addition to being essential for freerunning behavioural rhythms, *per* determines features of the temperature compensation mechanism.

Circadian systems persist in nature to anticipate daily environmental cycles. The effects of the *per* mutations on the patterns of *D. melanogaster* activity in light:dark (LD) cycles have been thoroughly investigated. Wild type *D. melanogaster* locomotor activity patterns tend to be bimodal, being most active at dawn and dusk, with the morning and

evening spurts of activity beginning a few hours before light-on and lights-off respectively (Hamblen-Coyle *et al.*, 1992; Petersen *et al.*, 1988). The per^{s} mutation advances the phase of the evening activity peak in LD, while the per^{Ll} mutation delays the onset of the evening peak (Dushay *et al.*, 1990; Hamblen-Coyle *et al.*, 1992; Konopka *et al.*, 1995). These expected effects on the phase of evening activity in LD contrast with the relative stability of the morning peak phase of these *per* mutants. The original arrhythmic allele per^{01} tends to be more active during lights-on than lights-off, with activity apparently changing in response to (rather than in anticipation of) the light cycle transitions (Hamblen-Coyle *et al.*, 1989; Petersen *et al.*, 1988). However, another arrhythmic allele, per^{04}, has LD activity patterns like wild type for a large proportion of flies tested (Hamblen-Coyle *et al.*, 1989), indicating that this mutant retains some level of *per* clock function.

1.3 Molecular biology of *per*

The molecular biology of *per* has been beset by problems. It was not until 1984 that the *per* gene was cloned; this was achieved by the Young group at Rockefeller University (Bargiello and Young, 1984a) and by the Hall and Rosbash groups at Brandeis University (Reddy *et al.*, 1984). The location of the clock gene was defined by the molecular mapping of chromosomal aberrations which interfered with *per* function. Previous cytogenetic analyses had placed the gene required for rhythmic behaviour in approximately polytene band 3B (Young and Judd, 1978; Smith and Konopka, 1981). Transcripts mapping within this interval were examined to determine which were specifically altered by chromosomal deletions that disrupted rhythmic behaviour. Two transcripts, a 4.5 kb and 0.9 kb species, apparently met this criterion (Bargiello and Young, 1984a), and therefore were considered to be the true *per* transcript. The 0.9 kb genomic sequence exhibited freerunning, circadian fluctuations in abundance (Reddy *et al.*, 1984), so it was also suggested that this RNA species was of special significance, both because of its abnormally low levels, and that its transcript was directly affected by per^{01} (Zehring *et al.*, 1984). However, it was later discovered that this transcript was neither necessary nor sufficient for rhythmicity (Hamblen *et al.*, 1986), but the 4.5 kb RNA did indeed play the central role. This was confirmed when genomic DNA fragments containing all or part of the 4.5 kb RNA transcription unit were transformed into per^{01} flies and tested for rescue of behavioural rhythmicity: all the genomic fragments containing the entire protein coding region of the 4.5 kb transcript

efficiently rescued the flies' circadian cycle (Bargiello *et al.*, 1984b; Zehring *et al.*, 1984; Hamblen *et al.*, 1986).

It was initially shown that alternative splicing of the primary transcript occurs giving rise to three species of mRNA (Citri *et al.*, 1987). The most abundant transcript, type A, has eight exons and contains a 1218 amino acid open reading frame, type B differs from the former by having an additional intron in the 3' untranslated region (3' UTR) as well as an unusually spliced intron which removes 96 amino acids within exon 5. The third, type C, has the first four introns removed, leaving a unique 107 amino acid sequence. Direct measurement of head RNA confirmed that type A is the most frequent *per* transcript, and that transcripts containing the type B intron in the 3'UTR are present but neither of the transcripts having type B specific intron in exon 5 nor type C transcripts were detected (reported in Hardin and Siwicki, 1995). This recent finding therefore suggests that only two *per* mRNA types exist and both encode identical PER polypeptides in the fly head.

The use of *per* mutant/wild type chimeric genes revealed that per^{L1} and per^{s} were due to missense mutations at amino acids 243 and 589, respectively, while per^{01} was found to be a stop codon at position 464 (Baylies *et al.*, 1987; Yu *et al.*, 1987a). Although the sequences surrounding the per^{s} and per^{L1} mutations were not recognised at the time, they have subsequently yielded some insights on PER function (see later). The sequence of the 127kD predicted PER protein initially revealed a similarity between a series of Threonine-Glycine (Thr-Gly) repeats of PER and sequences of Serine-Glycine repeats in some proteoglycan core proteins (Jackson *et al.*, 1986; Reddy *et al.*, 1986, Bargiello *et al.*, 1987). More recently biochemical studies have shown that PER is not a proteoglycan (Edery *et al.*, 1994b) and that the biochemical results of Reddy *et al.* (1986), and the physiological work of Bargiello *et al.* (1987) which suggested *per* was an intercellular coupling protein could not be replicated (Flint *et al.*, 1993). The Thr-Gly region is the most conspicuous within the PER protein, and secondary structure predictions suggest that the Thr-Gly region appears to act to separate PER into two domains (Costa *et al.*, 1991).

1.4 *per* expression

Expression of *per* is present in many cell types at various stages of development. In adult flies *per* is expressed in the nervous system and in several non-neuronal tissues including the gut and the gonads (Siwicki *et al.*, 1988; Saez and Young 1988; Liu *et al.*, 1988; Zerr *et al.*, 1990; Ewer *et al.*, 1992). Its expression in the lateral neurons of the brain

appears to be sufficient for generating locomotor activity rhythms, suggesting that these neurons are the site of the pacemaker (Ewer *et al.*, 1992; Frisch *et al.*, 1994). This is supported by recent evidence that, in certain strains of *per* transgenic flies which exhibit nearly normal circadian behavioural rhythms, the lateral neurons are the only cells in which PER protein is detectable (Frisch *et al.*, 1994). An extensive analysis of internally marked mosaics (Ewer *et al.*, 1992), found that the control of the circadian behaviour could not be mapped to a single discrete focus within the head. Weak rhythms were present in several mosaics with only small regions of per^{+} expression in putative glial cells in the ventral brain, suggesting that neuronal *per* expression is not necessary for circadian rhythms. Nonetheless, more robust rhythms were found in mosaics expressing per^{+} in lateral brain regions in a small group of neurons. The evidence suggests that rhythms resulting from lateral neuron *per* expression are stronger, especially since *disco* mutant flies which lack lateral neurons due to developmental defects, cannot generate freerunning rhythms even though robust and rhythmic glial cell *per* expression is observed (Zerr *et al.*, 1990; Helfrich-Forster and Homberg, 1993).

Probably the most significant aspect of *per* expression is that levels of *per* RNA and PER protein cycle during LD conditions and in freerunning conditions (Siwicki *et al.*, 1988; Hardin *et al.*, 1990; Zerr *et al.*, 1990). PER immunoreactivity was strongest near the end of the night phase (ZT21), and barely detectable near the end of the day (ZT12) in LD conditions and these fluctuations in abundance persisted in constant darkness (Zerr *et al.*, 1990). The abundance of *per* mRNA cycles 6-8 hours in advance of the protein in LD. *per* RNA accumulating during the middle of the day (ZT6) peaks early in the night (ZT15), both in LD and freerunning conditions (Hardin *et al.*, 1990). Protein and RNA cycles are altered predictably in per^{LI} and per^{s} flies in DD, indicating that RNA oscillations depend on PER protein, that is changes in *per* RNA levels give rise to changes in PER protein and that the protein feeds back to influence the timing of its own RNA's rhythm (Zerr *et al.*, 1990: Hardin *et al.*, 1990), establishing its own negative feedback loop. The observed time delay thus enables stable rhythms of expression, rather than the damping out of the feedback system to a 'steady state' of the clock. Consistent with this is the fact that *per* RNA does not cycle in per^{01} flies, which lack the functional PER protein (Siwicki *et al.*, 1988; Zerr *et al.*, 1990). This evidence makes *per* mRNA and PER protein excellent candidates for state variables, those quantities whose oscillations are essential to the maintenance of the circadian rhythm. This model would then predict that a transient increase in the amount of *per* expression at any one phase of the cycle should shift the clock to a new phase. In fact

such an effect was noted by Edery *et al.* (1994a), who used a heat inducible copy of *per* to mimic *per* expression. It would also predict that circadian oscillations should not occur in the absence of *per* mRNA oscillations, and indeed Frisch *et al.* (1994) demonstrated a dependence on the transcript oscillation of restoration of behavioural rhythmicity in per^{01} mutants rescued with *per* transgenes.

Oscillations of *per* RNA are known to be regulated at the transcriptional level, probably indirectly by PER protein (Hardin *et al.*, 1992). PER lacks a DNA binding domain but is expressed in the nuclei (Liu *et al.*, 1992), and it has been speculated that it regulates transcription by associating with a transcription factor (Huang *et al.*, 1993; Sehgal *et al.*, 1995). PER has a circadian pattern of phosphorylation, which may be involved in targeting it for degradation (Edery *et al.*, 1994b). The 6h lag between peaks of the *per* transcript and its product suggests some post transcriptional control. Curtin *et al.* (1995) showed that PER nuclear entry is temporally controlled as it accumulates in the cytoplasm before its translocation to the nucleus, and that the per^{L1} mutation delays nuclear entry in a temperature-dependent manner (Curtin *et al.*, 1995).

1.5 Partners of *per*

The biochemical basis for the delay in the *per* feedback system might involve another state variable of the clock, one of which may be the putative product of the *tim* gene. This gene was identified in a large scale autosomal screen for rhythm mutants (Sehgal *et al.*, 1994). Null tim^0 mutant flies completely lack rhythms in eclosion, locomotor activity, and molecular oscillations of *per* mRNA and PER protein, the latter of which is severely reduced in its overall level (Sehgal *et al.*, 1994; Price *et al.*, 1995). The nuclear localisation of a PER -β-galactosidase fusion protein is disrupted in *tim* flies (Sehgal *et al.*, 1994; Vosshall *et al.*, 1994), suggests that *tim* is a nuclear translocator. Thus, there is strong evidence that *tim* is essential for the *per* feedback loop and consequently for the function of the circadian clock, as the lack of oscillating *per* RNA in tim^0 flies is most probably caused by the absence of feedback by PER protein. The recent molecular characterisation of *tim* suggests how it might function in this process. The locus was positionally cloned, but the predicted 156kD TIM product was not related to any known protein (Myers *et al.*, 1995). However, *tim* RNA levels mimicked the *per* RNA oscillations in phase and amplitude (Sehgal *et al.*, 1995), and the cycling of *tim* mRNA ceased in per^{01} mutants , demonstrating

that both gene products are required to sustain their mutual interdependent molecular (and behavioural) oscillations.

Preliminary mapping experiments indicated that the PER -β-galactosidase fusion protein is responsive to *tim* only when it contains a region of PER that includes the PAS domain, a protein-protein interaction domain of around 250 amino acids which derives its name from the proteins that contain it: P for *per*, A for ARNT and AhR, the human aromatic hydrocarbon nuclear translocator receptor (Hoffman *et al.*, 1991) and the murine aryl hydrocarbon receptor (Burbach *et al.*, 1992), and S for the product of the neurogenesis gene *singleminded* (*sim*) (Crews *et al.*, 1988). The PER PAS domain interacts with these other PAS-containing proteins *in vitro* and also appears to interact with another domain of PER, called the C-domain (Huang *et al.*, 1995), when using the yeast two-hybrid system. The *tim* gene was also independently isolated using this system by Gekakis *et al.*, (1995), employing the PAS domain as the bait for the interaction. Experiments with glutathione-S-transferase fusion proteins *in vitro* confirmed a direct biochemical interaction between the PER and TIM proteins.

It was also shown in the same investigation that the per^{L1} mutation, which is within the PAS domain, may weaken the PER-TIM interaction. More specifically the per^{L1} mutation was shown to increase PER mediated intramolecular interactions with the C-domain and decrease intermolecular interactions (Huang *et al.*, 1995). The effects of per^{L1} on protein-protein interactions and nuclear transport (Curtin *et al.*, 1995) indicates that PER's interaction with other proteins affects its nuclear entry, which in turn probably determines the periodicity of the output rhythm. However, it is important to note that the timing of the nuclear gating or the length of the post-transcriptional lag, are not the only steps in the feedback loop that can affect the circadian period. Evidence of this is shown by the fact that the per^{s} mutation does not affect the timing of nuclear transport, yet it shortens its circadian rhythm by approximately 4-6 h (Konopka and Benzer, 1971; Curtin *et al.*, 1995). It is likely that the degradation of the PER^{s} protein occurs much earlier than wild type or it is exported from the nucleus earlier. Evidence that its nuclear function differs comes from the fact that PER levels decline more rapidly in per^{s} than in per^{L1} or in wild type (Zerr *et al.*, 1990; Edery *et al.*, 1994a). This also corresponds well to the per^{s} phase response curve (PRC) which has a reduced light insensitive period or 'deadzone', thereby shortening the period (Konopka, 1972; Winfree and Gordon, 1977).

A second element with which PER interacts is a protein kinase (or kinases), since PER has multiple phosphorylation sites, and the phosphorylation of PER is rhythmically

driven (Edery *et al.*, 1994b). The phosphorylation begins in the early part of subjective night, a time when the amounts of PER in the nucleus is on the rise, and appears greatest just before the decline of PER in late subjective night. It was suggested that the phosphorylation status of PER is an important determinant in the *Drosophila* clock's timekeeping mechanism. Furthermore circadian fluctuations in the abundance and phosphorylation of PER are suppressed in tim^{01} flies (Price *et al.*, 1995).

It is well known that light pulses can affect the phase of the circadian cycle, producing delays in the early evening, and advances during the latter half of the night (Pittendrigh, 1974; Saunders *et al.*, 1994). Lee *et al.* (1996) have recently showed that photic stimuli perturbed the timing of the PER protein and mRNA cycles in a manner consistent with the direction and magnitude of the phase shift. In addition the PER -TIM interaction *in vivo* was rapidly decreased by light. This disruption in the cytoplasm resulted in a delay of PER phosphorylation and nuclear entry, thus promoting phase delays in early night (ZT10 to ZT18), while disruption of the association in the nucleus was accompanied by an advance in PER phosphorylation and disappearance, suggesting phosphorylation could be the trigger for proteolysis (Edery *et al.*, 1994b) and accounting for the phase advances in late night (ZT19 to ZT04). These disruptions could be caused by the rapid degradation of TIM, but not PER, by light (Hunter-Ensor *et al.*, 1996), suggesting that TIM mediates light-induced resetting of the circadian clock. Light does not affect the levels of *per* and *tim* mRNA, indicating that the effects on TIM are postranscriptional (Hunter-Ensor *et al.*, 1996). In the early part of the night, a decrease in TIM levels by light is expected to delay the accumulation, and therefore the nuclear entry of PER - TIM heterodimers, thus delaying the events in the molecular loop. In the second half of the night, light would result in the premature disappearance of TIM , advancing the molecular rhythm (Hunter-Ensor *et al.*, 1996). This degradation by light is consistent with models for clock resetting which invoke photochemical destruction of clock components by light (Pittendrigh, 1993).

Implicit in the molecular feedback model is the existence of other loci whose transcription is also controlled by the PER - TIM complex but which function in the output pathways. One candidate is *Dreg-5* (Van Gelder *et al.*, 1995; 1996), whose transcript showed appreciable rhythms of expression in phase with *per* and *tim* mRNA cycles. These oscillations were predominantly dependent on PER and presumably TIM. However the accumulation of DREG-5 protein does not show the 6h delay observed for PER protein (Van Gelder, 1996), but peaks at ZT15, the same peak as its mRNA expression. However,

the protein levels did not decline as rapidly as mRNA levels, thus although *Dreg-5* and *per* are transcriptionally coupled, they have different post-transcriptional controls.

1.6 Clock genes in other organisms

In mammals behavioural circadian rhythms originate from a neural circadian clock located in the hypothalamic area of the brain known as the suprachiasmatic nucleus (SCN) (Kornhauser *et al.*, 1996). The genetic analysis of circadian rhythms in vertebrates began with the identification of a spontaneous mutant, *tau*, in hamsters (Ralph and Menaker, 1988), a single gene mutation that shortens the wheel running activity period. Like the *per* alleles this mutation is semidominant, homozygotes having a 20h period compared with a 22h period for heteroygotes. The *tau* mutation was thus formally similar to per^s in *D. melanogaster* , but one main difference is that homozygous individuals are unable to entrain to 12:12 LD cycles. Instead their periodicities drift into shorter rhythms, suggesting that the *tau* hamster's clock is unable to entrain (Ralph and Menaker, 1988). When a mutant *tau* SCN was transplanted into a wild-type lesioned host, the activity rhythm period was characteristic of the donor tissue's genotype (Ralph *et al.*, 1990). The reciprocal transplant worked as well, thereby confirming the SCN as the pacemaker.

In a search for genes that regulate circadian rhythms in mammals, the progeny of mice treated with *N*-ethyl-*N*-nitrosourea (ENU), were screened for semidominant circadian clock mutations (Vitaterna *et al.*, 1994). A mutation was isolated, *Clock*, which lengthens the intrinsic circadian period to 27h and abolishes the persistence of rhythmicity in homozygotes (Vitaterna *et al.*, 1994). Given the substantial genetic and mapping resources available for the mouse, it should be feasible to identify the nature of *Clock* by positional cloning.

Mutations have also been described in plants and bacteria, for example *Arabidopsis* (Millar *et al.*, 1995a) and cyanobacteria *Synechococcus* (Kondo *et al.*, 1993; 1994; Liu *et al.*, 1995). Reporter genes encoding bioluminescent luciferase markers were fused to circadian-regulated promoter fragments to provide non-invasive molecular markers for clock mutant screens. In *Arabidopsis*, the *CAB:luc* reporter system, in which the firefly luciferase gene is under the control of a CAB (chlorophyll a/b binding protein) promoter, has also been used to identify one of the photoreceptors in the circadian input pathway (Millar *et al.*, 1995b). These workers showed that the period of the *CAB:luc* luminescence rhythm was lengthened from 24 h to 30 h in DD in white light, whereas red or blue light

were able to confer 24 h periods similar to those in white light. Period shortening with red light was impaired in the phytochrome deficient mutant *hy1*, indicating that one or more of the phytochrome family mediates this response. In contrast period shortening by blue light was more efficient in *hy1* , indicating that a blue light receptor other than phytochrome mediates period shortening in blue light.

Ultradian rhythms, with a period shorter than 24h, are at least as widespread as circadian ones. Iwaski *et al.* (1995) have recently initiated the genetic detection of defaecation behaviour in *Caenorhabditis elegans*. This organism's defecation motor program runs on a 45*s* cycle, and this rhythm is temperature compensated. Mutations have been found to shorten or lengthen the cycle period and most of these disrupt temperature compensation suggesting this process is an integral part of the biological clock (Liu and Thomas 1994)

1.7 Other *per*-like sequences

Interesting sequences which have some similarities to *per* have also been found in other organisms. Synthetic oligonucleotides encoding three Thr-Gly repeats was found to be homologous to chicken, cat, mouse and human genomic DNA (Shin *et al.*, 1985). When using the mouse *per* repeat as a probe Matsui *et al.*(1993) found that a *per* repeat mRNA fluctuated in abundance in LD and in constant darkness in the SCN of the rat. Furthermore this mRNA expression was induced by light stimuli in a subjective night-specific manner, suggesting that the gene plays some role in the entrainment pathway of the mammalian circadian system (Matsui *et al.*, 1993). Similar sequence homology to the Thr-Gly repeats is found in higher plant nuclear DNA (rape and spinach) and in *Acetabularia* chloroplast DNA (Li-Weber *et al.*, 1987). It has also recently been found that two DNA fragments, 5.6 and 5.9 kb, that contain the Thr-Gly encoding ACNGGN repeat sequence, showed interesting associations with different circadian activity types of the blind mole rat (Ben-Shlomo *et al.*, 1996). This repeat also exists in the brain RNA of the mole rat, at higher quantities during the sleeping phase, suggesting that a protein composed of a poly Thr-Gly tract is involved in circadian behaviour (Ben-Shlomo *et al.*, 1996).

Remarkably the primary translation product of the *per* locus bears a limited similarity in the Thr-Gly region to the conceptual product of the *frequency* (*frq*) locus which determines conidiation (spore formation) circadian cycles in the bread mould *Neuropsora crassa* (McClung *et al.*, 1989). It was also seen that mutations of the *frq* gene

have the ability to shorten (*frq*^{1}), lengthen and lose partial capacity for temperature compensation (*frq*^{8}), and apparently obliterate (*frq*^{9}) the organism's circadian cycle of growth and spore formation. Therefore, *frq* appears to be a functional homologue of *per* (Feldman and Hoyle, 1973) and appears to be regulated in a similar way since *frq* RNA also displays circadian oscillations that are subject to negative feedback control by FRQ protein (Aronson *et al.*, 1994). It has also been elegantly demonstrated that the elimination of the molecular oscillation by constitutive overexpression of the putative FRQ protein results in loss of the overt rhythm (Aronson *et al.*, 1994). In addition, manipulation of the amount of *frq* gene product can alter the rhythm to a predicted phase. Crosthwaite *et al.*, (1995) found a rapid increase in *frq* expression upon exposure to brief pulses of white light, suggesting resetting happens extremely quickly after a light pulse, despite the fact the overt rhythm itself may require several days to reach a new steady state. It has also been suggested that resetting of the transcript rhythm of *frq* by transitory increases in *frq* mRNA in response to light could be the primary point of control of the clock mechanism (Loros, 1995). The *frq* gene in a related fungus, *Sordaria*, also has a similar region to the *N. Crassa* Thr-Gly repeat, although the dipeptides tend to also include Thr-Ala or Thr-Val which are conservative substitutions (Merrow and Dunlap, 1994).

1.8 PER sequence conservation

By comparing the published *per* sequences of *D. melanogaster* (Citri *et al.*, 1987), *D. virilis* and *D. pseudoobscura* (Colot *et al.*, 1988), *D. yakuba* (Thackeray and Kyriacou, 1990) and *A.pernyi* (Giant silkmoth, Reppert *et al.*, 1994), it is observed that this gene is a medley of relatively conserved sequences, blended between regions that cannot be aligned. The time elapsed since the common ancestors of Dipteran and Lepidoptera is about 200-250 million years (Myr), while the ancestors of the drosophilids compared, diverged from one another from 20-60 Myr ago (Hennig, 1981). Circadian rhythms are so widespread among animal and plant life, that they probably evolved early in the earth's history, irrespective of solar daylength in primitive times. Given this conservation of clock properties, and their ubiquitous nature, it would be thought that clock molecules would be conserved throughout the animal and perhaps the plant kingdom. Unfortunately it has turned out to be quite difficult to identify *per* homologues even outside drosophilids (Thackeray and Kyriacou, 1990; Colot *et al.*, 1988), and at present the best that has been

done is to isolate *per* in Muscids (*Musca domestica*, housefly, A. Piccin and C.P. Kyriacou, ms in prep.), Lepidoptera (*Antheraea pernyi*, giant silkmoth, *Hyalophora cecropia,* giant silkmoth and *Manduca sexta*, tobacco hornworm) , Orthoptera (*Periplanta americana*, American cockroach, Reppert *et al.*, 1994) and Hymenoptera (honeybee, Robinson, G. pers. comm.).

The *per* homologue from *A.pernyi* encodes an 849 amino acid protein, which shows highest identity (39%) with the PER protein of *D. virilis* (Reppert *et al.*, 1994). Regions of high similarity are in the amino terminus, the PAS region, and the region surrounding the per^{s} mutation in *D. melanogaster*. The *A. pernyi* gene also had a single Thr-Gly pair in the non-conserved region where the repetitive Thr-Gly region is found (Reppert *et al.*, 1994) in *Drosophila*. Moth *per* homologue mRNA expression exhibits a pronounced circadian cycle in adult heads, and PER antibodies show prominent variation of PER antigen staining in photoreceptor nuclei. PER protein from the giant silkmoth can also function in the *D. melanogaster* timing system (Levine *et al.*, 1995). Expression of the protein shortens circadian period in a dose dependent manner and restores pacemaker function to arrhythmic per^{01} mutants (Levine *et al.*, 1995). Surprisingly PER *Drosophila* antibodies in *M. domestica* showed only cytoplasmic staining of lateral neuronal cells across all circadian time points (A. Piccin, pers comm.), suggesting that PER in this Dipteran did not undergo nuclear expression and therefore it was maintained by a different mechanism than that found for *D. melanogaster*.

M. domestica PER is 1072 amino acids in length and stretches of similarity are found in all the conserved region first identified by Colot *et al.* (1988) in Drosophilid *per* genes. The PAS region is also well conserved (93.8%) and the structure of the Thr-Gly region is homogenous with all other Dipterans outside the *Drosophila* genus examined (Nielsen *et al.*, 1994), having two Thr-Gly pairs followed by one copy of a hexapeptide motif.

Within the conserved N-terminal region, but encoded downstream of the second PAS repeat lies the site of the per^{s} mutation, which shortens the circadian cycle in the fly (Konopka and Benzer, 1971). The region around this site is highly conserved (Colot *et al.*, 1988) and was used to generate an antibody to PER, called anti-S, which labelled a number of lateral brain neurons in the fly, as well as photoreceptor cells (Siwicki *et al.*, 1988; Zerr *et al.*, 1989). Furthermore, as discussed earlier the antigen appeared to cycle in abundance with a circadian period, giving the first clue as to the dynamic nature of the PER product (Siwicki *et al.*, 1988; Zerr *et al.*, 1990). Anti-S has been utilised to stain tissues in various

organisms in the hope that perhaps some of these labelled anatomical regions might include known circadian pacemakers, thereby indicating a conserved clock molecule. In these studies, an additional focus was whether the relevant antigen gave circadian changes in abundance, and whether the staining was nuclear at any circadian phase. It was therefore particularly exciting when Siwicki et al. (1989) reported that circadian pacemaker neurons in the eyes of the gastropods *Bulla* and *Aplysia* were labelled by anti-PER antibodies, and that in *Aplysia,* the antigen appeared to cycle in abundance in a circadian LD cycle. A 48 kD protein, recognised by the antibody was partially sequenced, and further polyclonal antibodies were raised against the relevant synthetic peptide (Strack and Jacklett, 1993). In turn, the new antiserum recognised a family of eye proteins, but on immunoblots, these did not show any evidence for circadian cycling (Strack and Jacklett, 1993). Immunohistochemistry revealed that several structures in the eye and head were labelled, but as yet, little can be concluded about the function of this protein.

The original anti-S antibody was also used with a carabid beetle, and labelled terminals within a putative photoreceptor in the optic lobe near the lamina (Fleissner *et al.*, 1993). However the relationship of this extraretinal organ to the circadian system is unknown, but the anti-S reactivity is intriguing, particularly as it is found in terminals, suggesting a secreted protein. The antibody also labelled a number of neural structures which receive inputs from the pineal and the retina in *Xenopus* (Garcia - Fernandez *et al.*, 1994), but again, these results are currently meaningless with respect to circadian biology. Finally, anti-S also labelled neurons in the suprachiasmatic nucleus (SCN) of the rat, the site of the rodent circadian pacemaker (Siwicki *et al.*, 1992). Immunoblots identified three bands, one of which, 160-170 kD showed a daily rhythm in the SCN, peaking in the middle of the dark phase, but this did not persist under freerunning conditions (Rosewell *et al.*, 1994). As yet no gene or protein sequence corresponding to the rodent antigen has been generated. Thus the possibility that a conserved PER-like motif in the circadian system and pacemakers of these organisms is related to clock function, remains little more than speculation at present.

The comparative analysis of *per* genes and their patterns of expression, will eventually reveal whether the negative feedback model based on the cycling of both *per* and *tim*, is conserved within the insect class and beyond.

1.9 The Threonine-Glycine repeat

Of the non-conserved regions, the one that carries the Dipteran Threonine-Glycine (Thr-Gly) repeat has received the most attention. In *D. melanogaster* there are about 20 pairs of these alternating residues (Colot *et al.*, 1988; Yu *et al.*, 1987b), but even within drosophilids, this is not conserved (Thackeray and Kyriacou, 1990; Peixoto *et al.*, 1992; 1993b; Nielsen *et al.*, 1994). High levels of length polymorphism in the Thr-Gly encoding repeat region of the *per* gene were observed in natural populations and laboratory strains of *D. melanogaster* (Yu *et al.*, 1987b; Costa *et al.*, 1991; Costa *et al.*, 1992). The 14, 17, 18, 20, 21, and 23 Thr-Gly length variants were found, interestingly, with the exception of the $(Thr\text{-}Gly)_{18}$ length allele, all the other Thr-Gly alleles differ by multiples of $(Thr\text{-}Gly)_3$. Comparisons of the Thr-Gly sequences within the *melanogaster* subgroup showed a rapid evolution in this region by a combination of slippage like events (see Dover, 1989) and point mutations (Peixoto *et al.*, 1992). However the variation in length of the Thr-Gly region among the eight species of the subgroup analysed is surprisingly small, given the potential for length differences that was observed in *D. melanogaster* alone (Costa *et al.*, 1991). These observations might suggest some kind of restraint is acting on the length of this region. This view is supported since in spite of the fact that the size range within and between species is similar, the amino acid composition between species does vary (Peixoto *et al.*, 1992), revealing that sufficient time has elapsed for significant evolutionary change to occur within the *melanogaster* subgroup.

Dipteran outside the *Drosophila* have two pairs of Thr-Gly's (Fig. 1.1. *Beris vallata, Syritta pipiens, Musca domestica, Loxocera albisata, Rhagoletis completa,* , *Zaprionus tuberculatus*, aswell as within *Drosophila, Drosophila pictcornis* and *Drosophila pictiventris*) and this appears to be the ancestral state (Nielsen *et al.*, 1994), when compared to the Drosophilidae, in which the region is much larger and extremely variable, both in size and composition. In *D. pseudoobscura* a length polymorphism is observed in a five amino acid degenerate repeat (Colot *et al.*, 1988; Costa *et al.*, 1991). The *D. pseudoobscura* pentapeptide has undergone expansion to 30-35 copies (Colot *et al.*, 1988; Costa *et al.*, 1991; Nielsen *et al.*, 1994), but *D. pseudoobscura* also carries about 10 copies of the Thr-Gly encoding hexamer. This pentapeptide originates from a hexapeptide sequence found in larger non-drosophilid flies, which itself appears to have been initially generated by an ancient expansion of the Thr-Gly repeat, Fig. 1.1 (Nielsen *et al.*, 1994).

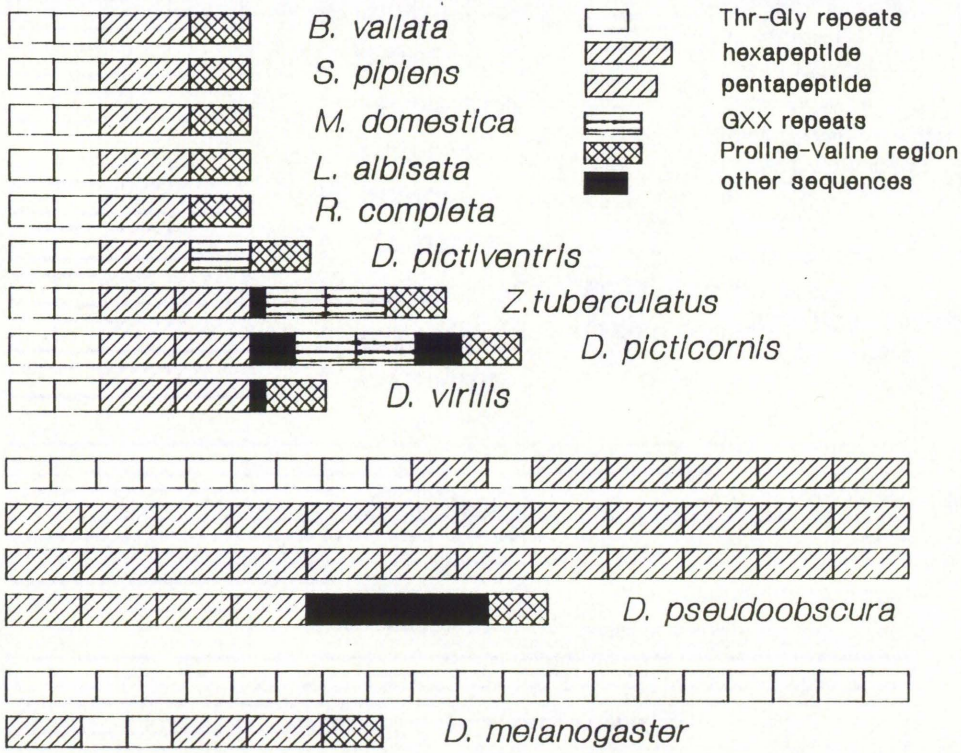


Figure 1.1 Cartoon showing the expansion of the repetitive region within the diptera (redrawn from Nielsen *et al.*, 1994)

Theoretical models (Gray and Jeffreys, 1991; Harding *et al.*, 1992) indicate that the initial expansion of a repetitive region requires at least two tandem repeats. After a certain period of expansion followed by a stabilisation, in which a large number of repeats are maintained, there is a reduction to a small copy number followed by a collapse to a single stable copy. For example, in great apes and Old and New World monkeys, the MS32 noncoding hypervariable minisatellite appears stable and monomorphic, having only three to four repeat units. In humans, however, the same minisatellite shows extreme variation in copy number (Gray and Jeffreys, 1991). If this hypothesis of events is correct then other insect orders would have maintained Thr-Gly pairs but do not have the larger repeat units. This has found to be the case in *A. pernyii* and other insect orders as described previously (Reppert *et al.*, 1994), where for instance the giant silkmoth has one Thr-Gly pair followed by a Pro-Gly pair. In spite of the differences between *D. melanogaster* and *D. pseudoobscura* in the amino acid sequence of the repeats, the predicted secondary structures suggest evolutionary and mechanistic constraints on the PER protein of the two species (Costa *et al.*, 1991), with the Thr-Gly region dividing the conceptual protein into two domains.

Peixoto *et al.* (1993b) sequenced and compared the Thr-Gly region from several different *Drosophila* species belonging to various groups within the *Drosophila* and *Sophophora* subgenera. Analysis of the data suggested that changes in the length of this variable region might be associated with amino acid replacements in the more conserved flanking DNA (Peixoto *et al.*, 1993b). Further support for the view that length changes in the repeat region are associated with 'coevolutionary' changes in the flanking regions was also observed by Nielsen *et al.* (1994). The critical experiments to determine whether the length of the Thr-Gly region has coevolved with the immediate flanking region required the analysis of hybrid *per* genes between *D. melanogaster* and *D. pseudoobscura* where the exact positions of the chimeric junctions involving the Thr-Gly repeat and the flanking sequences were manipulated thereby breaking up the proposed 'co-evolution' (Peixoto, PhD thesis 1993a, I. Townson and J. M. Henessey pers. comm.). This is described in more detail in chapter 5.

1.10 Population biology of *per*

As described previously the Thr-Gly repeat shows extensive variation in *D. melanogaster* laboratory strains (Yu *et al.*, 1987b) and natural populations (Costa *et al.*,

1991; 1992), with European flies carrying from 14 to 23 pairs of uninterrupted Thr-Gly's. From the many alternative phylogenetic relationships that could exist between them only a few have actually occurred (Rosato *et al.*, 1996), primarily by a number deletion and duplication events (Costa *et al.*, 1991; Rosato *et al.*, 1996). Also, it has been postulated that the $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ alleles are the most ancient, and that from them all the other length alleles can be derived (Rosato *et al.*, 1996).

A population survey of European flies revealed that the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length variants made up approximately 90% of all individuals typed, with the $(Thr\text{-}Gly)_{23}$ comprising about 9% and the $(Thr\text{-}Gly)_{14}$ the other 1% (Costa *et al.*, 1992). Other rarer variants were found carrying 18 and 21 Thr-Gly pairs, but the most striking feature of this investigation was the very significant latitudinal cline that was found for both major length variants (Costa *et al.*, 1992). The $(Thr\text{-}Gly)_{17}$ allele was found to be more frequent in southern Europe, whereas the $(Thr\text{-}Gly)_{20}$ allele predominates in northern Europe. This evidence raised the possibility of natural selection maintaining this length polymorphism, although historical factors, for example, population admixture, are also known to create clines. It was thought that no simple relationship existed between single environmental factors and the genetic variation, since no correlations were found with a number of environmental variables, including temperature and rainfall (Costa *et al.*, 1992).

D. simulans, the sympatric sibling species of *D. melanogaster*, also shows length polymorphism in the Thr-Gly repeat, but this is largely limited to 23, 24 and 25 dipeptide pairs in natural European populations (Rosato *et al.*, 1994). However, unlike *D. melanogaster* where only one flanking amino acid substitution has been found (Rosato *et al.*, 1996), each length variant of *D. simulans* carries a unique flanking amino acid haplotype which is in perfect linkage disequilibrium (Rosato *et al.*, 1994). No clinal distribution was found for these length alleles and the relative frequencies are similar in all eight European locations sampled. Classical neutrality tests suggest that the variation observed is consistent with a balancing selection scenario, so why does *D. simulans* not show a cline? I will return to this question in chapter 4.

Structural studies and molecular dynamics simulations of Thr-Gly dipeptides reveal that a $(Thr\text{-}Gly)_3$ peptide appears to form a stable beta-turn and may represent a conformational monomer (Castiglione-Morelli *et al.*, 1995; see also Ishida *et al.*, 1994). This is an interesting observation bearing in mind that the major Thr-Gly length variants found in natural populations of *D. melanogaster* differ by a $(Thr\text{-}Gly)_3$ periodicity, i.e. the

$(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ series, suggesting that these are the more stable repeat lengths (Costa *et al.*, 1992).

1.11 *per* and speciation ?

Petersen *et al.*(1988) reported that species-specific behavioural instructions are transferred by the *D. pseudoobscura per* coding region when it was transformed into per^{01} *D. melanogaster* mutants. The transformants had shorter periods and hence resembled those of *D. pseudoobscura* as did the eduction profiles of locomotor activity in LD entrainment, which showed one major peak of activity. This species-specific circadian locomotor behaviour pattern has been mapped to the N-terminal half of *per* using interspecific chimeric genes between *D. melanogaster* and *D. pseudoobscura* (J.M. Hennessy and C.P. Kyriacou, pers comm.).

Molecular transfer of a species-specific behaviour has also been demonstrated using *D. simulans* and *D. melanogaster* (Wheeler *et al.*, 1991). *Drosophila* males modulate their song interpulse intervals (IPI) during courtship by extension and vibration of their wings. These intermittent bursts of tone pulses and hums carry information which enhances the females receptivity for mating (Cowling and Burnet, 1981; Kyriacou and Hall, 1982; Kyriacou and Hall, 1986; Wheeler *et al.*, 1991). Species-specific pulse information is found within the IPI's which averages 30-40ms in *D. melanogaster* and 45-55ms in *D. simulans* (Cowling and Burnet, 1981; Wheeler *et al.*, 1991), and by the modulation of the mean IPI, ranging from 50-65s in *D. melanogaster* and 30-40s in *D. simulans* (Kyriacou and Hall, 1980; 1982). A cloned copy of the *D. simulans per* (13.2s-TGs), which corresponded to the 13.2 kb *D. melanogaster per* DNA (13.2m-TGm) used in the restoration of rhythms to per^{01} *D. melanogaster* flies (Citri *et al.*, 1987; Yu *et al.*, 1987b; Petersen *et al.*, 1988; Hamblen-Coyle *et al.*, 1989) was introduced into the *D. melanogaster* per^{01} genome. Transformations were also performed using chimeric constructs of the 13.2 kb *D. melanogaster* fragment which carried the Thr-Gly region of *D. simulans* (13.2m-TGs), the reciprocal construct was also made (13.2s-TGm). All constructs rescued per^{01} locomotor activity arrhythmicity showing that they can provide basic *per* function. The mean song cycles of 13.2m-TGm males gave the expected *D. melanogaster* periods (Kyriacou and Hall, 1980; 1986; 1989; Kyriacou *et al.*, 1990b) of about 55s. Males carrying a transduced

13.2s-TGs *per* gave mean period lengths of 35s characteristic of *D. simulans* (Kyriacou and Hall, 1980:1986). Males carrying the chimeric construct with the *D. simulans* Thr-Gly region, 13.2m-TGs, also gave *D. simulans* like song rhythms of 35s cycles, whereas the reciprocal chimerics (13.2s-TGm) had song periods ranging from 55-58s.

Irrespective of the difference in Thr-Gly length between the two species (a difference of 4 Thr-Gly pairs) the only differences observed are found downstream of the Thr-Gly repeat where four species-specific amino acid substitutions are found, and by inference one or more of these changes are most likely to account for the species-specific song rhythm (Wheeler *et al.*, 1991). Deletion of the Thr-Gly region plus a few amino acids was also shown to effect the male courtship song rhythm (Yu *et al.*, 1987b), dramatically shortening this ultradian rhythm from 60s in the wild type to approximately 40s in the *per* deleted transformants.

D. melanogaster females do not discriminate in mating tests between strains of *D. melanogaster* which have been genetically transformed with the *per* gene of either *D. melanogaster* or *D. simulans* or with wild type *D. melanogaster* (Ritchie and Kyriacou, 1994). This contradicted the playback experiments which indicated that *D. melanogaster* or *D. simulans* females mate more readily when stimulated with song containing the appropriate IPI pattern (Kyriacou and Hall, 1982: 1986; Greenacre *et al.*, 1993). This therefore suggests that *per* may make only a small contribution to total premating isolation between these sibling species (Ritchie and Kyriacou, 1994) or suggests that the genetically transformed individuals may not have been exhibiting robust rhythmic song patterns.

The *per* locus has been proposed as a 'speciation gene' because of the species - specific influences of both these circadian and ultradian behaviours (Coyne, 1992). These characteristics of *per* biology might suggest that the relevant domains that determine these interspecific differences may not be evolving neutrally, but be subject to selection. A particularly suitable scenario would be 'selective sweeps', where a preferred variant 'sweeps' through a population, leaving behind a population carrying the selected allele with all its linked nucleotide variation.

Kliman and Hey (1993) sequenced a number of *per* alleles in *D. melanogaster* and three of its sibling species and performed various neutrality tests on a 1.9 kb region which spanned from the second exon to just short of the Thr-Gly domain. They found no compelling evidence for directional selection acting on this region and likewise no significant departure from neutral expectations in the observed frequencies of patterns of polymorphism with any of the species. This was in agreement with a study by Begun and

Aquadro (1991), which also provided no evidence for selection at the *per* locus in *D. melanogaster* and *D. simulans*.

Another study of nucleotide variation at the *per* locus of the *D. athabasca* complex suggested that *per* may not play a role in mating isolation between these groups, and that they have diverged very recently (Ford *et al.*, 1994). This also implies that the mating isolation differences and song differences (Yoon, PhD thesis 1991) came about very quickly (Yoon and Aquadro, 1994). Hypothesised phylogenetic relationships between the three semispecies of *D. athabasca*, based on mitochondrial DNA data, suggest that patterns of sympatry and allopatry strongly implicate the action of selection in the rapid evolution of behavioural isolation for this complex (Yoon and Aquadro, 1994).

An examination of neutrality of the Thr-Gly region in *D. melanogaster* would be welcomed, however, this is not directly possible, since mutational mechanisms other than base substitutions have occurred, and these contradict the assumptions of the infinite-site model on which these tests are based (Watterson, 1975; Dover, 1987). Fortunately the HKA (Hudson *et al.*, 1987) and Tajima tests (1989) of neutrality could be applied to the Thr-Gly region of *D. simulans* (Rosato *et al.*, 1994), but this was only possible because the three major length alleles were in perfect linkage disequilibrium with specific flanking haplotypes. Thus the HKA and Tajima tests could be focused on the adjacent sequences (Rosato *et al.*, 1994). In this instance, the HKA (Hudson *et al.*, 1987) test compared within-species polymorphism and between-species divergence in two regions of the genome, *per* and *rosy*, with the sequences flanking the Thr-Gly repeats in *D. simulans*. No significant departures from neutral expectations were obtained. In the Tajima (Tajima, 1989) test the number of segregating sites (i.e., polymorphic) was compared to the average number of nucleotide differences between the pairs of sequences, and this resulted in a positive and highly significant D statistic. This shows a significant departure from a strictly neutral model of molecular evolution.

In *D. melanogaster* an examination of neutrality is more difficult since 13 isolength alleles, including 7 Thr-Gly length classes, have been identified (Costa *et al.*, 1991; 1992; Peixoto *et al.*, 1992), and the same Thr-Gly length variant can be associated with more than one flanking haplotype (Rosato *et al.*, 1996), and this precludes the use of the classic neutrality tests. However Rosato *et al.*(1996) extended the analysis to the Thr-Gly domain by using different statistical analyses of associated linkage disequilibrium, and a pattern of variation consistent with weak selection was found.

1.12 Project outline

It is now well established that a fundamental mechanism controlling the circadian pacemaker in *D. melanogaster* is under auto-regulatory or cross-regulatory control of the critical clock loci, *tim* and *per* (Van Gelder and Krasnow, 1996). These and other transcription translation oscillators (TTO's) are units of the circadian program (Pittendrigh, 1996). One of the fundamental principles of the output pathways of these TTO's that we now take for granted, temperature compensation, was discovered in 1952 by Pittendrigh, albeit by comparing adult emergence in an outhouse and a pressure cooker anchored in a creek to sustain a constant low temperature (reported in Pittendrigh, 1993). Of course this was followed up in the laboratory and led to the publication of the temperature independent eclosion rhythm of *D. pseudoobscura* (Pittendrigh, 1954). Therefore, it seems appropriate in the year that clockwatchers lost this giant, that I present here a report which includes an insight into the natural variation of temperature compensation in *D. melanogaster.* This study starts in chapter 3 with a sequence analysis of the Thr-Gly length variants used in my studies. Chapter 4 describes the geographical analysis of the Thr-Gly length polymorphism in natural *D. melanogaster* populations collected in Australia, and the various statistical analyses thereof. In chapter 5 the freerunning locomotor activity rhythms of European and North African Thr-Gly length variants are presented, and discussed with reference to the cline previously found in these populations (Costa *et al.*, 1992). Chapter 6 takes two of the populations studied in the chapter 5 and investigates their distributions of activity and thereby their energy expenditure during 12:12 LD cycles. The locomotor activity observations performed in DD on Australian Thr-Gly length variants originating from different latitudes are presented in chapter 7, and finally chapter 8 describes a simple population cage experiment set up at two different temperatures with natural Thr-Gly length variants.

CHAPTER 2

2.1 Stock maintenance

Fly stocks were reared on sugar/agar medium (4.63 g of sucrose, 4.63 g of live yeast, 0.71 g of agar and 0.2 g of Nipagen in 100 ml of water), in either one-third pint milk bottles or glass vials (10 cm x 2.2 cm). The flies were kept in temperature controlled rooms at either 18±1°C or 25±1°C and in light-dark cycles of 12 hours (LD12:12) with lights on/off at 09:00/21:00 h. The fly strains and populations used in this work are described in the relevant chapters.

2.2 Locomotor activity experiments.

The locomotor activity experiments were carried out using a locomotor activity event recorder produced by BIODATA Ltd. (Manchester, U.K.). Each fly was loaded into a cylindrical glass tube (8 x 0.3 cm) containing 2 to 3 cm of sugar/agar medium at one end, sealed with tape so that the food would not dry out. Tissue paper or cotton wool was used to seal the other end. Each tube was placed between an infra-red light emitter and detector, which recorded an event every time the fly crossed and broke the light beam. Infra-red wavelengths were used because flies are insensitive to this region of the spectrum (Bertholf, 1932; Kyriacou and Burnet, 1979). The number of events registered in every half-hour time window (bin) was then fed directly into a Tandon 386 computer (IBM PC compatible).

The males and virgin females used in the locomotor activity experiments were raised at 25°C in light/dark cycles (LD12:12) with lights on/off at 09:00/21:00 h. The freerunning experiments ran for seven days using the following protocol. Two days *before* data collection, flies one to seven days old were loaded in the tubes and placed in an incubator which was set to the desired temperature and the same light/dark cycle in which the flies had been raised. On the following day the incubator was set automatically to darkness beginning from 21:00 h and 18 hours later (at 15:00 h), data collection began for a period of 7 days.

Before analysis, the data were edited in the following ways. Flies which died during the experiment before the end of the fifth day of data collection were excluded from analysis. Also excluded from analysis were flies from channels considered hypersensitive,

that is, which exhibited unrealistic counts, from 500 to 1000 events, in many half-hour bins. Flies which died within the last two days of data collection were included in the analysis after deleting from the activity record approximately the last 12 hours of data, to the nearest half day, before the cessation of activity. Very occasionally an otherwise 'normal' record showed an isolated bin with very high counts. For example, in one case a bin with 3007 events was observed for a fly which showed for its entire record less than 60 events per bin during its peak activity phase. This apparently spurious data point was corrected by simple interpolation, that is, by replacing the value of that bin with the average value of its two neighbouring bins. As a rule any bin with more than 100 events, which also gave 10 times more events than its interpolated estimate was corrected in this way. These errors were probably caused by a 'bug' in the software which could not be resolved. However, their appearance was so rare that they did not pose a significant problem. Similar problems have been described for other activity monitors by Hamblen-Coyle *et al.* (1992).

The periodicity in the locomotor activity is sometimes evident in the raw data or can be visualised by inspection of the 'actogram'. Figures 2.1a and b illustrate, respectively, a graph of the activity data and an actogram of a rhythmic *D. melanogaster* carrying a $(Thr\text{-}Gly)_{20}$ length allele. In the actogram, the data are double plotted so that the events of day one and day 2 are plotted in real time on the first horizontal line. Below this line are plotted the events from day 2 and 3, and below this the data from days 3 and 4 and so on. A rough estimate of the period can be obtained by tracing a straight line to connect the times of activity 'onset' or 'offset'. However this is very often misleading and more rigorous ways to detect rhythmicity use statistical techniques such as autocorrelation or spectral analysis (Dowse and Ringo, 1989).

In the work presented here, the periodicity of locomotor activity for each fly was initially determined by using autocorrelation analysis (Diggle, 1990) available in the SPSS/PC+ Version 2.0 software package. Briefly, the number of events obtained at two time points separated by a certain time lag are correlated. Thus with a lag of one time bin, bin 1 data is correlated with bin 2, bin 2 with bin 3, etc... If there is a 24 h cycle in the data and each bin represents a 30 min window, bin 1 and bin 49, bin 2 and bin 50, bin 3 and bin 51, etc., should produce the highest correlation coefficient. The correlation coefficients are then plotted against each 'lag' (1,2,3 ... etc.) in a correlogram which extends to 116 lags (58 h). Figure 2.1c shows the correlogram of the activity data presented in Fig. 2.1a. As can be seen the autocorrelation function shows a peak at 24 h (lag 48) and a second peak close to its harmonic at 48.5 h. This represents an extremely conservative and robust estimation of

Figure 2.1a Histogram of activity data plotted over 7 days in 30 minute bins

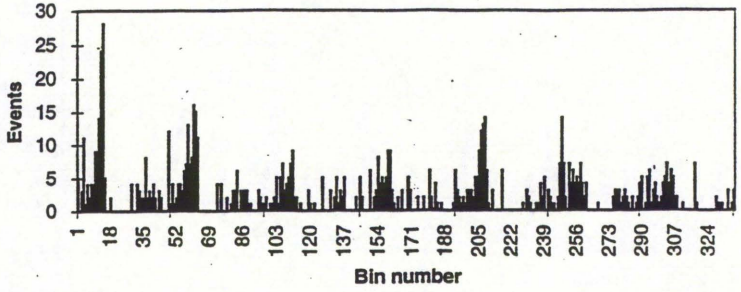


Figure 2.1b Actogram, double plotted data showing day 1 and day 2 on first line followed by day 2 and day 3 on second etc.

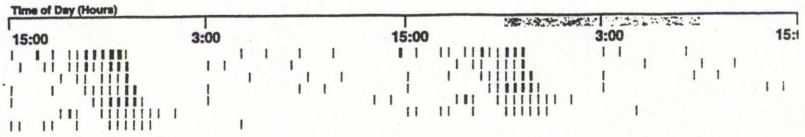


Figure 2.1c Correlogram, plot of correlation coefficients for each time lag up to 58h in 30 minute bins, horizontal lines represent the approximate 95% confidence intervals.

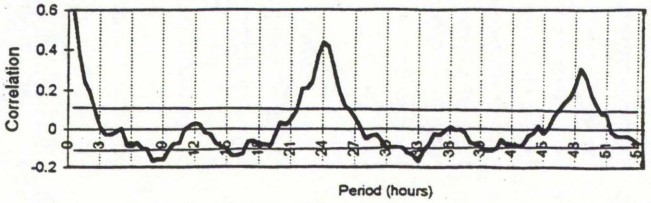
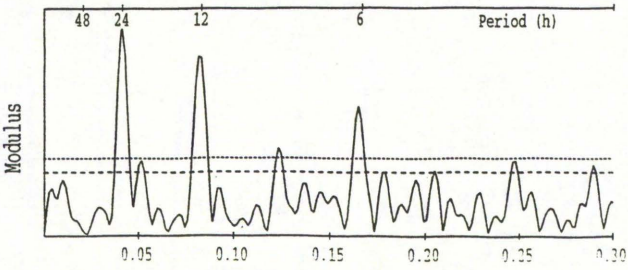


Figure 2.1d Spectrogram showing the periods which best describe the data plotted against the strength of such a period on the Y axis. The horizontal lines represent the 95%(--) and 99% (..)confidence intervals respectively.



rhythmicity in a digital signal (Chatfield, 1980). The horizontal lines in the correlogram are the approximate 95% confidence limits ($2/\sqrt{N}$; where N is the number of bins). For a fly to be considered rhythmic, the highest peak in the profile within the range 15-40 h is taken as its period of activity, so long as its peak lay above the 95% confidence limit. The decision to consider a fly rhythmic or not was also made by taking into account the overall pattern of its correlogram, which in a rhythmic fly will be sinusoidal. This implies some degree of subjectivity in distinguishing an arrhythmic fly from a fly showing weak rhythms with a 'noisy' background. Moreover, the so called 'aperiodic' per^{01} mutants are not completely arrhythmic and display single or multiple ultradian rhythms in the 4-22 h range (Dowse *et al.*, 1987). Long and weak rhythms (between 29 and 40 h) have also been reported for these flies (Ewer *et al.*, 1990). Therefore the classification of flies into rhythmic or arrhythmic was performed 'blind' without prior knowledge of its genotype, as was the assignment of the period of activity.

A spectral analysis was also performed using a fourier transformation based on the CLEAN algorithm (Roberts *et al.*, 1987; Kyriacou and Hall, 1989). Spectral analysis fits sine and cosine waves to the signal and assigns a wave pattern, of period, p, which best fits the data (Fig 2.1d). The spectrogram in Fig 2.1d gives the periods which best describe the signal, and their power plotted on the Y axis (modulus). The highest relevant peak was taken as the period. However, if a large peak was found in the 12 h domain this period was doubled giving a period which was nearly always concordant with the circadian period observed with the correlogram. In Fig 2.1c the 24 h rhythm is stronger than the 12 h peak and this is mirrored in the spectrogram (Fig. 2.1d). A 12 hour peak is found at predominantly higher temperatures when the activity occurs approximately 12 h apart at dawn and dusk (see chapter 6). The data for each fly were randomised and the spectral analysis repeated. This was done 100 times, and the modulus values for each run at each frequency were ordered from lowest to highest. The 95th and 99th values in ascending order represent the approximate 95 and 99% confidence limits, based on this Monte Carlo simulation of random data. To be judged 'significant' a peak had to be greater than the 99% confidence limit. In Fig 2.1d the dashed horizontal line represents the 95% confidence limit, and the dotted line the 99% confidence limit. The flies circadian periods were entered into a database, and the relevant information was periodically extracted on an Excel spreadsheet for the statistical analysis which followed. Only flies which were rhythmic for both their spectral and autocorrelation analyses were considered rhythmic.

2.3 Molecular techniques

2.3.1 Fly genomic DNA extraction for PCR

Fly DNA used in the PCR reactions was prepared using the method of Gloor and Engels (1990). Single males were ground in 50 µl of squishing buffer (10 mM Tris-HCl pH 8.2, 1 mM EDTA, 25 mM NaCl, 200 µg/ml Proteinase K), left for 30 min at 37°C and then heated above 95°C for 5 min. 1 µl of genomic DNA was used for each 10 µl of PCR reaction.

2.3.2 PCR Amplification, Gel Electrophoresis and Direct DNA Sequencing

PCR was carried out according to Jeffreys *et al.* (1988b). Each 10 µl of reaction contained 1.35 µl of PCR mix, 1 µl of each 10 µM primer and 0.15 µl of Taq (5 units/µl) plus DNA solution and water. Two drops of mineral oil were added to prevent evaporation during the PCR. The recipe to prepare a 7.4X PCR mix was the following (final concentration in the PCR reaction between brackets):

167.0 µl 2 M Tris-HCl pH 8.8 (67 mM)
83.0 µl 1 M ammonium sulphate (16.6 mM)
33.5µl 1 M $MgCl_2$ (6.7 mM)
3.6 µl 14.38M 2-mercaptoethanol (10 mM)
3.4 µl 10 mM EDTA pH 8.0 (6.7 mM)
75.0 µl 100 mM dATP (1.5 mM)
75.0 µl 100 mM dCTP (1.5 mM)
75.0 µl 100 mM dGTP (1.5 mM)
75.0 µl 100 mM dTTP (1.5 mM)
85.0 µl 10 mg/ml DNase free BSA (170 µg/ml)

Taq polymerase from different sources such as Taq polymerase from Advanced Biotechnologies and Taq polymerase from Amersham were used. Reactions were cycled for 1 min at 95°C, 1 min at 65°C , and 1 min at 70°C for 30 cycles in a DNA thermal cycler (either Techne Programmable Dri- Block PHC-1 , a Perkin-Elmer-Cetus DNA thermal cycler, or Biometra TRIO - thermoblock.).

To determine the frequencies of the different threonine-glycine (Thr-Gly) length variants, PCR products were electrophoresed through a 2.5% to 3.5% low melting point

NuSieve (GTG) agarose gel. TBE buffer (0.045M Tris-Borate, 0.001M EDTA) was used, and PCR amplified DNAs from previously characterized length alleles, such as the (Thr-Gly)$_{14}$, (Thr-Gly)$_{17\text{-}2}$, (Thr-Gly)$_{17}$, (Thr-Gly)$_{20}$, (Thr-Gly)$_{21}$ and (Thr-Gly)$_{23}$ laboratory stocks of *D. melanogaster* (Yu *et al.*, 1987b), were used as markers (e.g. see Fig. 3.1, chapter3).

The length variants defined were further characterised by co-amplification of each DNA, with a previously sequenced DNA from an isolength Thr-Gly allele. Co-amplification was made using a 1:1 ratio of the two DNA's. If co-amplification produced a heteroduplex this indicated a difference in the DNA sequence of the variant versus the isolength standard (see Fig. 8.1 chapter 8). Also for heterozygous females an extra heteroduplex band also forms from the two different length variant strands originating from the two X chromosomes, compared to only one in the male. These were compared with co-amplification of different Thr-Gly length variants.

Double-stranded direct DNA sequencing of PCR products was carried out according to Bachmann *et al.* (1990) with the dideoxy chain-termination method (Sanger *et al.*, 1977) using the Sequenase version 2.0 kit from United States Biochemical. The PCR product to be sequenced was purified in a 1% low melting point agarose gel (GTG NuSieve) and recovered by phenol and double phenol-chloroform extraction followed by ethanol precipitation. Alternatively, the PCR fragments to be sequenced were purified from normal agarose gels using the QIAEX II agarose gel extraction kit (QIAGEN Inc.) or the QIAquick-spin PCR purification kit(250) (QIAGEN Inc.). Automated sequencing was also carried out using a modified version of the ABI Prism Dye Terminator Cycle Sequencing Ready Reaction Kit followed by double water saturated phenol-chloroform extraction and ethanol precipitation. In this instance a quick and easy PCR product isopropanol purification was also used (Dr. M. Poidinger, Dept Microbiology, University of Western Australia, Nedlands, 6009, Australia) for the automated sequencing.

2.3.3 Primers

Table 2.1 shows a list of the primers used in PCR , direct DNA sequencing and automated sequencing. The first column gives the primer names. The second column shows the sequence and the third column gives the primer positions based on the *D. melanogaster per* gene sequence (Citri *et al.*, 1987). The approximate primer positions in relation to the *D. melanogaster* protein sequence are illustrated in Fig. 2.2.

TABLE 2.1 List of primers used for PCR, direct and automated sequencing. The primer positions refer to the *D. melanogaster per* gene sequence (Citri *et al.*, 1987).

primer	sequence	position
5'per8	5'-ATACACATGAGCAGTGTGAC-3'	5066-5085
5'per2	5'-AACTATAACGAGAACCTGCT-3'	4874-4893
5'PTGME	5'-GTCCACGAGGGCAGCGGGGG-3'	5009-5028
3'per2	5'-TTCTCCATCTCGTCGTTGTG-3'	5336-5355
3'per4	5'-TCCATCTCGTCGTTGTGCTT-3'	5333-5352
3'PTGME	5'- GCGCGACTCGCGGTGCTTCT-3'	5365-5386

2.3.5 Analysis of the DNA sequences.

The sequences obtained with the ABI automated sequencer were analysed using the Autoassembler and Factura programmes on an Apple Macintosh.

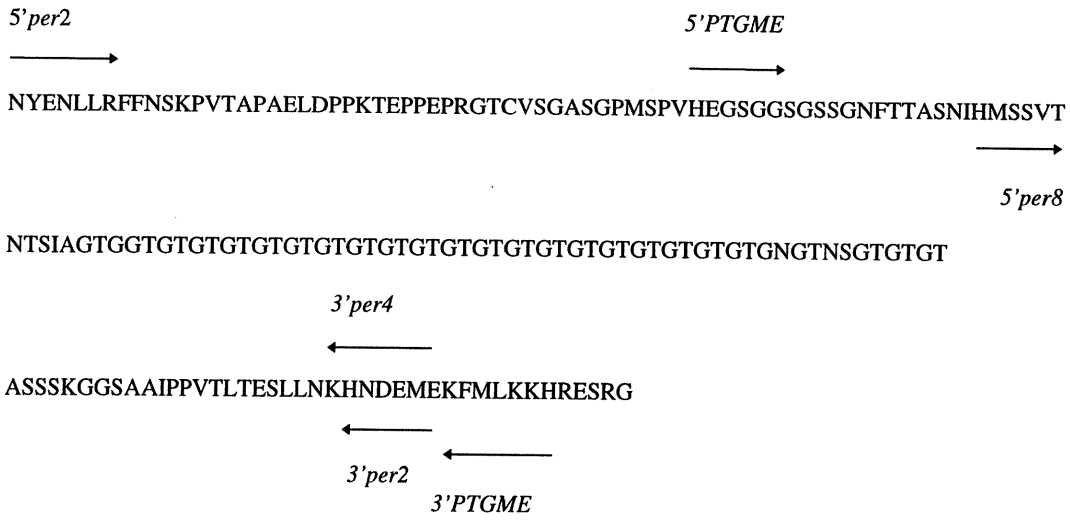


Figure 2.2 Approximate positions of the primers listed in Table 2.1. for PCR, direct and automated sequencing.

CHAPTER 3

DNA SEQUENCE ANALYSIS OF THE LENGTH POLYMORPHISM IN THE Thr-Gly REPEAT REGION OF THE *per* GENE IN *Drosophila melanogaster.*

3.1 INTRODUCTION

In this chapter the DNA sequences encoding the Thr-Gly region of the *per* gene of *D. melanogaster* are examined. As mentioned in chapter 1, this region of *per* is known to show length variation between different laboratory and natural populations (Yu *et al.*, 1987b., Costa *et al.*, 1991, 1992). Following the geographical analysis of the length polymorphism in *D. melanogaster* of Australian populations which will be presented in chapter 4, new Thr-Gly length variants were found. These sequences plus those of the Thr-Gly length variants used in the behavioural analyses (chapters 5 and 7) are presented here. In addition one new European length variant, NW5, a $(Thr\text{-}Gly)_{24}$ length allele, was also found, as were some new length alleles from Kenya. These unpublished data are presented here and compared to those published in Costa *et al.* (1991) Peixoto (PhD thesis 1993a) and Rosato *et al* (1996). Apart from the general interest value of coding repeats to molecular evolution, a practical reason for examining these sequences is to see whether isolength alleles are indeed identical in sequence, both in the repetitive and flanking regions. If not, then it may be possible to derive the alleles from different ancestors. Given that the Thr-Gly region is likely to be in linkage disequilibrium with flanking sequences, it is important for the understanding of the behavioural data, to ascertain the origins of these variants in case of behavioural heterogeneity within an isolength variant.

3.2 METHODS

Single fly DNA extraction, PCR, direct DNA sequencing and automated DNA sequencing were carried out as described in chapter 2. The sequences presented here are taken from flies collected from different localities in Europe (Rodolfo Costa), Australia (Avis James and Linda Partridge)and Kenya. The samples from Pietrastornina (PI), Merano (ME) in Italy, and Bordeaux (BX) in France were obtained in October 1989 while the samples from Casablanca (CA) in North Africa, Leiden (LE) in the Netherlands and Rethimnon (RE) in Crete were collected in September-October 1989. Flies from Cognac (CO) in France, San Mateo (SM) and Albaida (AL) in Spain were collected in October-November 1991. The samples from Conselve (CON) in Italy were collected in October 1992 and individuals were also used from a sample collected in Platanistassa (PLAT), Cyprus in September 1992. The samples from Australia were all collected in February 1993. A population from North Wooton, Somerset, UK was collected in October 1994 (gift from Pete Corish), and a sample was also collected from Kenya during October and November 1995 (NGMI, KESI and K, gift from Stephen Andersson). For more details see chapter 4 and Costa *et al.* (1991 and 1992). All the individuals under study were males either immediately frozen after capture, or derived from isofemale lines immediately established after collection of the sample.

3.3 RESULTS

3.3.1 Thr-Gly length polymorphism in natural European and North African *D. melanogaster* populations

Several of the sequences compared here were isolated in the initial characterisation of the Thr-Gly length polymorphism (Costa *et al.*, 1991) and in the subsequent geographical analysis of natural *D. melanogaster* populations (Costa *et al.*, 1992). A number of alleles from European populations were typed and sequenced. It was observed that the most

common alleles showed mobility corresponding to 17, 20 and 23 Thr-Gly pairs (see chapter 4 and Costa *et al.*, 1991). Rare variants having mobilities corresponding to 14, 15, 18 and 21 Thr-Gly repeats were also found. A number of alleles from other populations were also sequenced following the behavioural analysis of the polymorphism across Europe and North Africa (chapter 5). An example of separation of PCR-amplified Thr-Gly variants obtained by agarose gel electrophoresis is shown in Fig. 3.1.

Figs. 3.2 and 3.3 show the DNA sequences of the different Thr-Gly length variants. All Thr-Gly alleles detected in the individuals whose behaviour was analysed (chapter 5, 6 and 7) were sequenced. These sequences were also compared to the previously published sequences CS, OR and CH, cloned respectively from the laboratory strains Canton-S (Jackson *et al.*, 1986), Oregon-R (Citri *et al.*, 1987) and Chieti (Yu *et al.*, 1987b). Sequences published in Costa *et al.* (1991), included ME3 (Merano, Italy), BX1 (Bordeaux, France) and PI1 (Pietrastornina, Italy). Sequences presented in Peixoto (PhD thesis 1993a) and Rosato *et al* (1996), are AL40, AL25 and AL56 (Albaida, Spain), CO109 and CO40 (Cognac, France), and SM3 and SM24 (San Mateo, Spain). PLAT 2b, PLAT 6 and PLAT 8 (Platanistassa, Cyprus) are published in Rosato *et al.* (1996).

A new length variant was found in the North Wooton population, NW5, a $(Thr\text{-}Gly)_{24}$ length repeat. Previous to this, the $(Thr\text{-}Gly)_{23}$ was thought to be the longest length allele segregating in natural or laboratory strains. Therefore the observation of a longer length allele was interesting from a phylogenetic and behavioural point of view (see chapter 5). The possible events leading up to the derivation of the $(Thr\text{-}Gly)_{24}$ will be discussed later in this chapter.

3.3.2 The sequences flanking the Thr-Gly repeat in European and North African populations.

Fig. 3.2 shows the alignment of the sequences in the regions flanking the uninterrupted repeat. The sequences are compared to that of the longest Thr-Gly length variant, $(Thr\text{-}Gly)_{24}$, NW5, and the nucleotide positions are numbered at the bottom of the figure with the predicted protein translation given above the DNA sequences. The

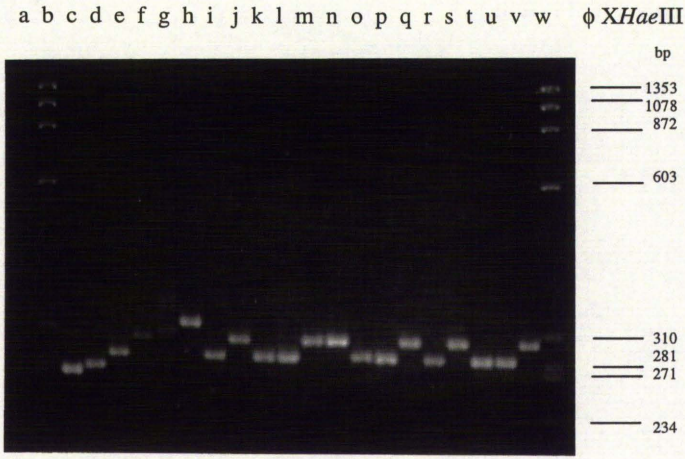


Figure 3.1 PCR-amplified *D. melanogaster* Thr-Gly length variants separated on a 3.5% low melting point gel (Nusieve). Amplification is with *5'per8* and *3'per2* , see Table 2.1 and Fig. 2.2. Lane a is a blank; lane b and w are ϕ X174 RF/*Hae*III size markers; lane c, $(Thr\text{-}Gly)_{14}$; lane d, $(Thr\text{-}Gly)_{17\text{-}2}$; lane e, $(Thr\text{-}Gly)_{17}$; lane f, $(Thr\text{-}Gly)_{20}$; lane g, $(Thr\text{-}Gly)_{21}$; lane h, $(Thr\text{-}Gly)_{23}$; lanes i, k, l, o, p, r, t and u are $(Thr\text{-}Gly)_{17}$ length variants; lanes j, m, n, q, s and v are $(Thr\text{-}Gly)_{20}$ length variants.

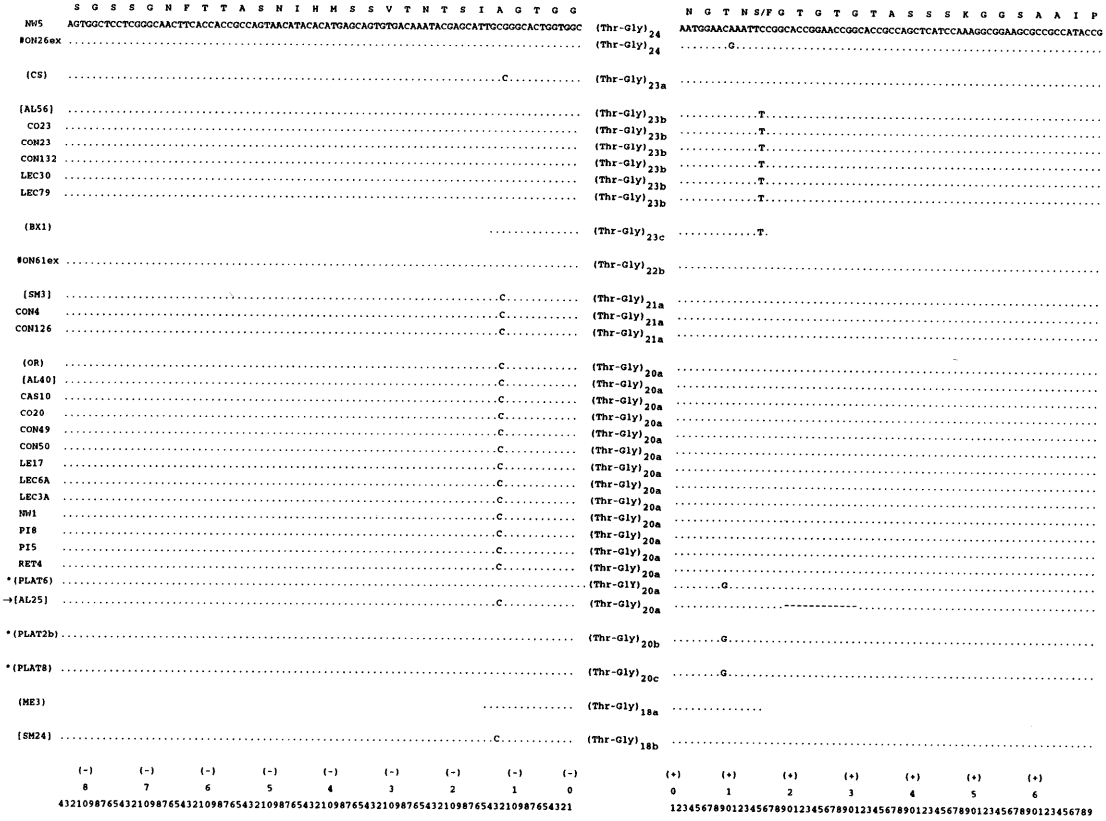


Figure 3.2. DNA sequence of the regions flanking the uninterrupted Thr-Gly repeat of European and North African flies (see following page for full legend).

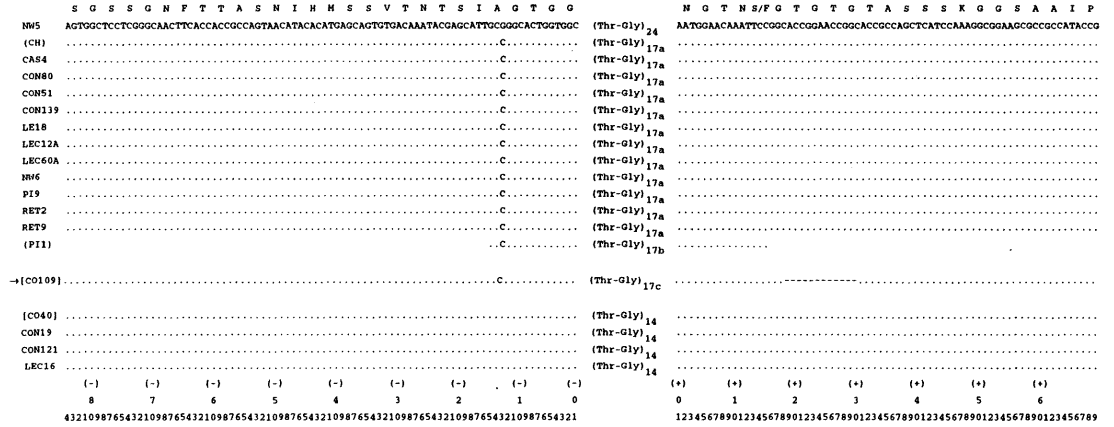


Figure 3.2. continued. DNA sequence of the regions flanking the uninterrupted Thr-Gly repeat of European and North African flies. The dots indicate conserved positions, the dashes indicate deletions. The amino acid sequence is shown at the top of the figure and the single amino acid substitution (S/F) is indicated. The numbers at the bottom are used to label each nucleotide position. Published sequences are reported in brackets ( ). Note that the BX1, ME3 and PI1 sequences shown here are 3 bp longer, both in the 5' and the 3' regions flanking the main repeat, than in Costa *et al.*, (1991); however the sequences for the entire flanking regions were not available. The sequences described by Peixoto (PhD thesis 1993a) are shown in parenthesis [ ]. The sequences described by Rosato *et al.*,(1996) are prefixed with *. African sequences prefixed with # are personal communication (C. Aquadro). The sequences marked with → indicate a Thr-Gly repeat lacking the downstream (Thr-Gly)$_2$ repeat. BX = Bordeaux (France), CAS = Casablanca (Morocco), CH = Chieti (Yu *et al.*, 1987b), CO = Cognac (France), CON = Conselve (Italy), CS = Canton S (Jackson *et al.*, 1986), LE= Leiden (Netherlands), LEC = Lecce (Italy), ME = Merano (Italy), PI = Pietrastornina (Italy), PLAT = Platanistassa (Cyprus), NW = North Wooton (U.K.), OR = Oregon-R (Citri *et al.*, 1987), SM = San Mateo (Spain), RET = Rethimnon (Crete).

predicted number of Thr-Gly pairs in the main uninterrupted repeat according to the mobility of the PCR products is shown between the 5' and the 3' sequences. It was previously shown that PCR fragments generated using external primers such as *5'per8* and *3'per2* (see Chapter 2) can give a misleading impression of the number of Thr-Gly pairs in the uninterrupted repeat (Peixoto, PhD Thesis 1993a). The arrow (→) marks two of the previously published cases where the sequencing revealed that the length variation in *D. melanogaster* was not only due to changes in the number of Thr-Gly pairs in the uninterrupted repeat, but also due to the existence of a deletion of two Thr-Gly encoding repeats downstream of the main repeat. As a consequence of this deletion, PCR fragments generated from these two length variants using external primers, produce fragments which have a mobility corresponding to alleles with 18 and 15 Thr-Gly pairs respectively. However, no new lines of this type were discovered in the sequencing of the European length alleles, suggesting that this is an extremely rare deletion event (Rosato *et al.*, 1994, 1996).

After sequencing the regions immediately flanking the Thr-Gly repeat alleles (Fig. 3.2), no flanking polymorphisms were seen when the European variants were compared to the previously published common length variants. A single nonconserved replacement substitution was observed when comparing European $(Thr\text{-}Gly)_{23}$ length alleles with the Canton-S strain, also a $(Thr\text{-}Gly)_{23}$ length variant. Otherwise this C → T substitution which is responsible for a serine to phenylalanine (S → F) replacement is in perfect linkage with the $(Thr\text{-}Gly)_{23b}$ and $(Thr\text{-}Gly)_{23c}$ alleles, as previously observed by Rosato *et al.*, 1996.

3.3.3 Sequence analysis of the perfect Thr-Gly encoding repeats of natural European and North African populations.

Fig. 3.3 shows the DNA sequences encoding the uninterrupted repeat. The sequences are aligned according to the length of the perfect uninterrupted repeat and the arrows are used to highlight different Thr-Gly encoding cassettes. At present there are 8 known length classes corresponding to 14, 17, 18, 20, 21, 22, 23 and 24 Thr-Gly pairs. Apart from the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{24}$ length variants all the other length classes have

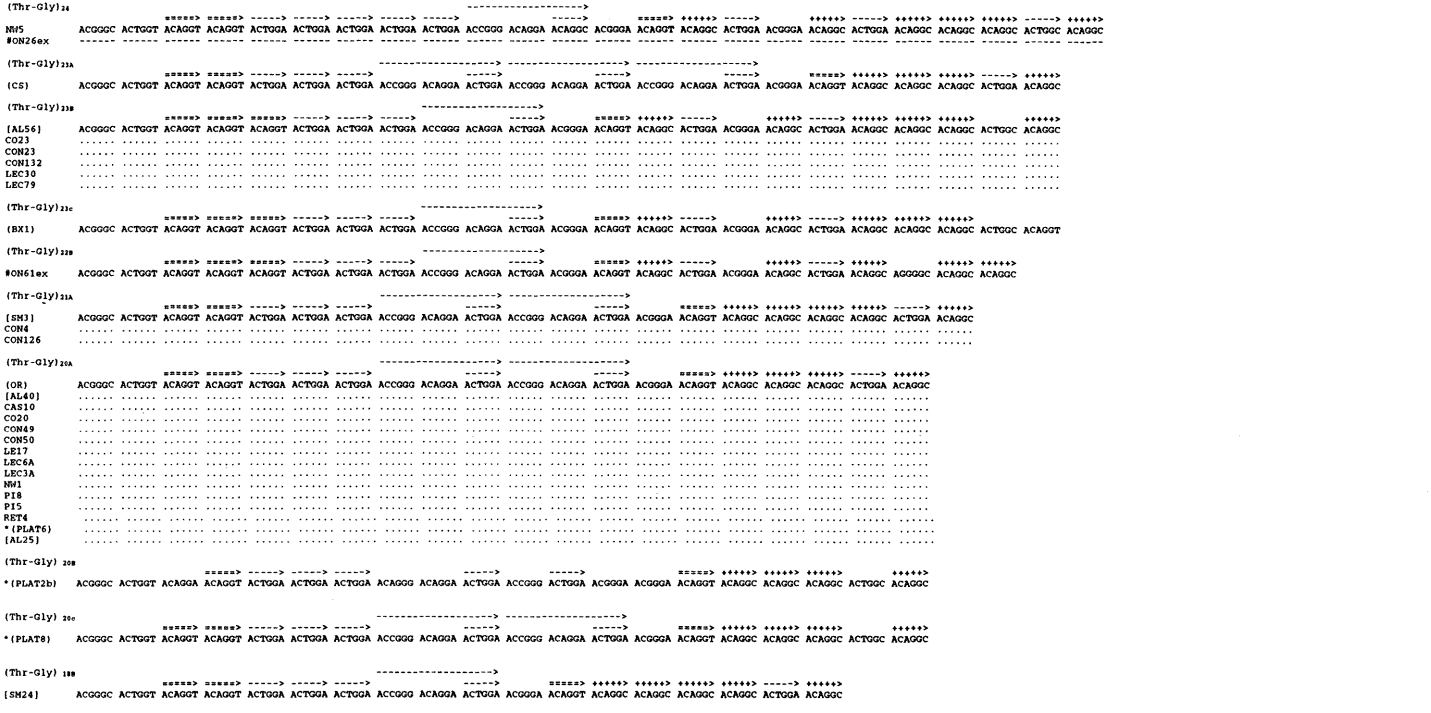


Figure 3.3. Nucleotide sequences of the perfect repeat of the *D. melanogaster* Thr-Gly region in European natural populations (see following page for full legend).

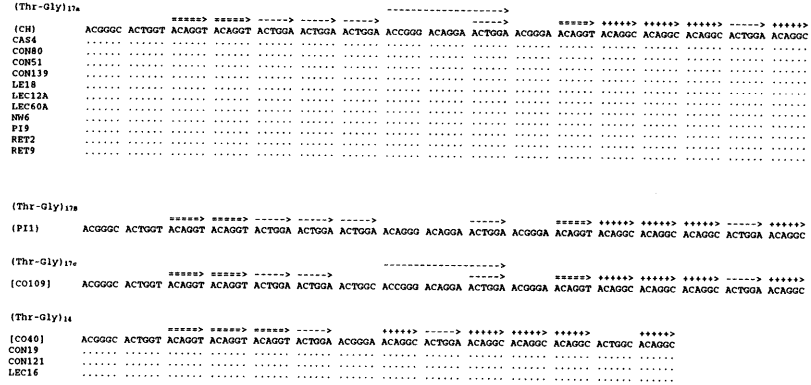


Figure 3.3. continued. Nucleotide sequences of the perfect repeat of the *D. melanogaster* Thr-Gly region of European and North African flies. The dots indicate conserved nucleotide positions. The arrows highlight similarities among the alleles (see text). The sequences are ordered on the basis of the length of the region and according to the different alleles considered. Published sequences are reported in brackets ( ) (Costa *et al.*, 1991) , those prefixed with * are from Rosato *et al*(1996) while the sequences described by Peixoto (PhD thesis 1993a) are shown in parentheses [ ]. Sequences prefixed with # are personal communication (C.Aquadro). BX = Bordeaux (France), CAS = Casablanca (Morocco), CH = Chieti (Yu *et al.*, 1987b), CO = Cognac (France), CON = Conselve (Italy), CS = Canton S (Jackson *et al.*, 1986), LE= Leiden (Netherlands), LEC = Lecce (Italy), ME = Merano (Italy), PI = Pietrastornina (Italy), PLAT = Platanistassa (Cyprus), NW = North Wooton (U.K.), OR = Oregon-R (Citri *et al.*, 1987), SM = San Mateo (Spain), RET = Rethimnon (Crete).

isolength alleles which have silent substitutions in the third positions of the Thr-Gly encoding ACN GGN hexamers. Since all the sequences reported here for the European and North African length alleles are identical to the published data, their phylogenetic relevance and evolutionary derivation will be discussed in concert with the Australian Thr-Gly length alleles.

3.3.4 Thr-Gly length polymorphism in natural Australian *D. melanogaster* populations

The sequences examined here were originally found in the identification of the Thr-Gly length polymorphism in the Australian populations and the following spatial analysis (Chapter 4). Several length alleles were typed and sequenced, and as before with the European length alleles the majority of the variants showed mobilities corresponding to $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length repeats. However other lengths were also observed which corresponded to $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{18}$, $(Thr\text{-}Gly)_{21}$, $(Thr\text{-}Gly)_{22}$, and $(Thr\text{-}Gly)_{24}$ length repeats.

3.3.5 The sequences flanking the Thr-Gly repeat in Australian populations.

All alleles whose carriers were examined behaviourally (Chapter 5, 6 and 7) were sequenced as were a few flies from Kenya, see Fig. 3.4. These sequences are also compared to the previously published sequences Canton-S (Jackson *et al.*, 1986), Oregon-R (Citri *et al.*, 1987) and Chieti (Yu *et al.*, 1987b). The other published sequences shown are as referenced in Section 3.3.1.

Six nucleotide substitutions and one deletion are observed in Fig. 3.4. A synonymous change in relation to the NW5 sequence is observed in an Alanine codon (G → C), position 5101 in Citri *et al.*, 1987, position -13 in Fig. 3.4. At this position C is associated with all $(Thr\text{-}Gly)_{17}$, most of the $(Thr\text{-}Gly)_{20a}$, the $(Thr\text{-}Gly)_{18b}$, $(Thr\text{-}Gly)_{18c}$, $(Thr\text{-}Gly)_{21a}$, most of the $(Thr\text{-}Gly)_{23a}$, and $(Thr\text{-}Gly)_{22a}$ variants. In contrast the $(Thr\text{-}Gly)_{24}$, $(Thr\text{-}Gly)_{23b}$, $(Thr\text{-}Gly)_{23c}$, all but one of the $(Thr\text{-}Gly)_{22}$, the $(Thr\text{-}Gly)_{18a}$ and $(Thr\text{-}Gly)_{14}$ alleles are always in linkage with the G polymorphism at position 5101, as are the

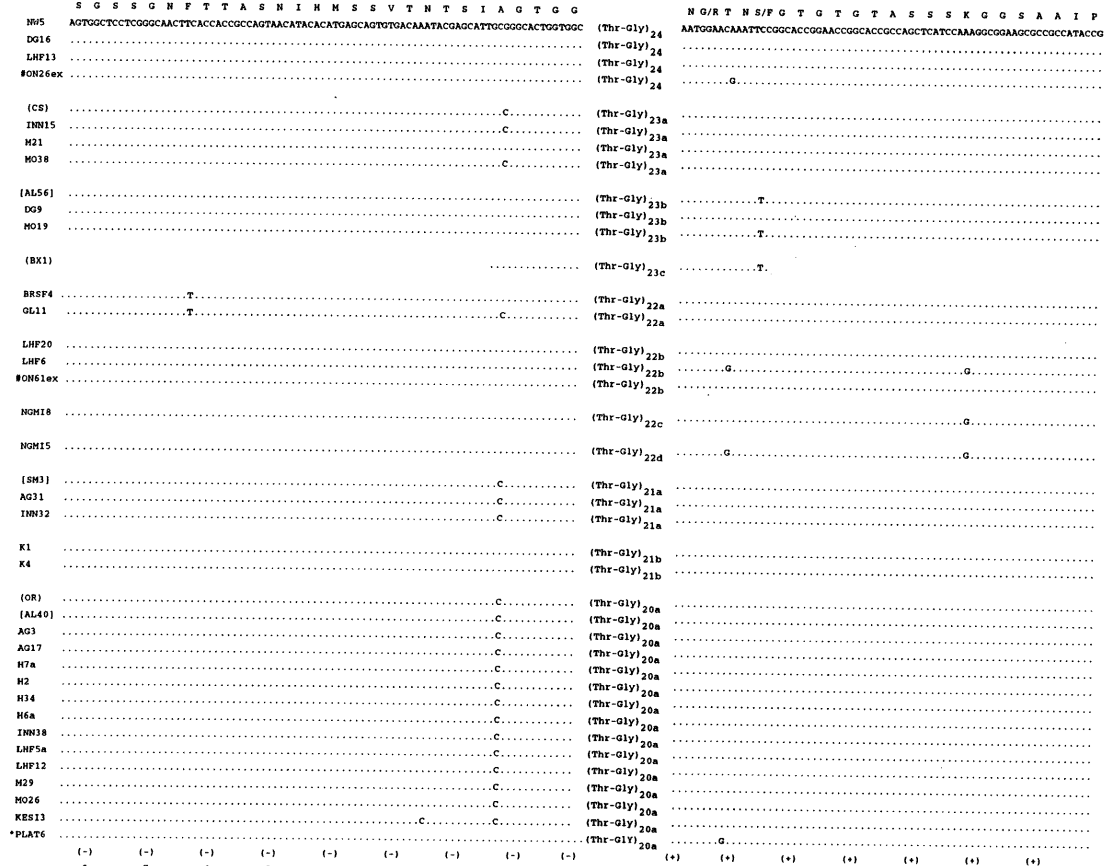


Figure 3.4. DNA sequence of the regions flanking the uninterrupted Thr-Gly repeat of Australian and Kenyan flies (see following page for full legend).

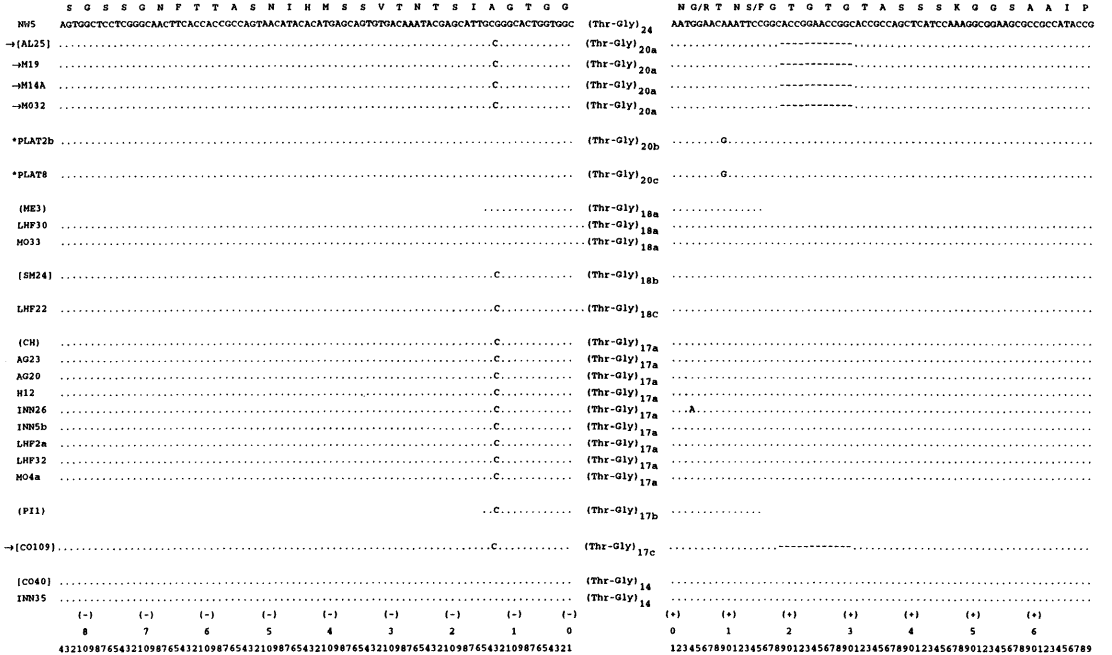


Figure3.4. continued. DNA sequence of the regions flanking the uninterrupted Thr-Gly repeat of Australian and Kenyan flies. The dots indicate conserved positions, the dashes indicate deletions. The amino acid sequence is shown at the top of the figure and the only two amino acid substitutions (S/F and G/R) are indicated. The numbers at the bottom label each nucleotide position. Published sequences are reported in brackets ( ). Note that the BX1, ME3 and PI1 sequences shown here are 3 bp longer, both in the 5' and the 3' regions flanking the main repeat, than in Costa *et al.*, (1991); however the sequences for the entire flanking regions were not available. The sequences described by Peixoto (PhD thesis 1993a) are shown in parentheses [ ]. The sequences described by Rosato *et al.*,(1996) are prefixed with *. Sequences prefixed with # are personal communications (C.Aquadro). The sequences marked with an arrow (→) indicate a Thr-Gly repeat lacking the downstream $(Thr\text{-}Gly)_2$ repeat. AG = Yeppoon (Queensland) ,AL = Albaida (Spain), BRFS = Bowen (Queensland) , BX = Bordeaux (France), CH = Chieti (Yu *et al.*, 1987b), CO = Cognac (France), CON = Conselve (Italy), CS = Canton S (Jackson *et al.*, 1986) , DG = S. Brisbane, (New South Wales), GL = Hervey Bay (Queensland) , H = Melbourne (Victoria), INN = Cairns (Queensland) , K = Kiambu, Nairobi (Kenya) , KESI = Kericho (Kenya), LHF = Yeppoon (Queensland) , M =Melbourne (Victoria) , ME = Merano (Italy), MO = Cairns (Queensland), NGMI = Nguruman (Kenya) , NW = North Wooton (U.K.), OR = Oregon-R (Citri *et al.*, 1987), PI = Pietrastornina (Italy), PLAT = Platanistassa (Cyprus), SM = San Mateo (Spain).

three different $(Thr\text{-}Gly)_{20}$ length alleles from PLAT2b, PLAT 6 and PLAT 8. Rosato *et al.*, (1996) noted that the only $(Thr\text{-}Gly)_{20}$ alleles associated with the G(5101) come from Cyprus, and because this association is relatively rare, it infers a close phylogenetic relationship among all the Thr-Gly sequences sharing the G(5101) marker. This hypothesis is upheld by another polymorphism 3' of the main repeat. In position 5242 (position +9 in Fig. 3.4) a G nucleotide is associated with the same $(Thr\text{-}Gly)_{20}$ sequences from Platanistassa (PLAT2b, PLAT 6 and PLAT 8). However, this is not exclusive since a $(Thr\text{-}Gly)_{22b}$ (LHF6, Yeppoon, Queensland) and a $(Thr\text{-}Gly)_{22d}$ allele (NGMI5, Nuguruman, Kenya) also have this polymorphism, G(5242), as well as the G(5101) marker. All other length alleles are in linkage disequilibrium with the A nucleotide at the position 5242. This unusual combination of markers and length alleles from very few lines could tentatively be suggested to be due to founder effects, since it is very unlikely that six separate but identical mutations occurred, particularly three in the same population.

A silent (C/T) polymorphism in position 5060 (position -64 in Fig. 3.4), where T is associated with two $(Thr\text{-}Gly)_{22a}$ variants, BRSF4 (Bowen, Queensland) and GL11 (Hervey Bay, Queensland) is also shown. A third silent 5' flanking polymorphism only appears in one $(Thr\text{-}Gly)_{20a}$ allele, KESI3 (Kericho, Kenya), where there is a T $\rightarrow$ C substitution in position 5101 (position -25 in Fig. 3.4). A non - synonymous substitution (C $\rightarrow$ T, position 5247 +14 , phenylalanine, in Fig. 3.4) was thought to be in perfect linkage disequilibrium with the $(Thr\text{-}Gly)_{23b}$ and $(Thr\text{-}Gly)_{23c}$ alleles (Rosato *et al.*, 1996). However DG9 (S. Brisbane, New South Wales) a $(Thr\text{-}Gly)_{23b}$ allele has the C nucleotide at this position (serine), so it is only accurate to say that the rare $(Thr\text{-}Gly)_{23c}$ variant is associated with this marker. This polymorphic site involves a non-conservative amino acid change (S $\rightarrow$ F) in a region implicated in the control of the interspecific differences in the courtship song cycle (Wheeler *et al.*, 1991). Previously the (C/T) polymorphism at position 5247 was thought to be the only non conservative substitution in the Thr- Gly region, however a second non-synonymous change has been found in a single $(Thr\text{-}Gly)_{17a}$ allele, INN 26 (Cairns, Queensland). This replacement substitution is responsible for a glycine to arginine (G $\rightarrow$ R) amino acid change in position 5237 (position +4 in Fig. 3.4).

The sixth nucleotide substitution at position 5281 (position +48, A → G) is found in three lines, a $(Thr\text{-}Gly)_{22b}$ (LHF6, Yeppoon, Queensland), the $(Thr\text{-}Gly)_{22c}$ (NGMI8, Nguruman, Kenya) and the $(Thr\text{-}Gly)_{22d}$ variants (NGMI5, Nguruman, Kenya). This would suggest a recent mutation associated with only the $(Thr\text{-}Gly)_{22}$ length alleles, but one not in complete linkage disequilibrium as other isolength alleles have the A nucleotide at this position.

Finally, the sequences of the alleles AL25 (Albaida, Spain), M19, M14A (Melbourne, Victoria), MO32 (Cairns, Queensland), and CO109 (Cognac, France) show that they are not variants having respectively 18 and 15 Thr-Gly pairs as suggested by the length of the PCR products. Instead they represent uninterrupted $(Thr\text{-}Gly)_{20a}$ and $(Thr\text{-}Gly)_{17c}$ variants which share the same downstream deletion of two Thr-Gly pairs. Because this deletion appears to be rare in *D. melanogaster* (see Costa *et al* 1992), the fact that these different length variants ('pseudo $(Thr\text{-}Gly)_{18}$' and 'pseudo $(Thr\text{-}Gly)_{15}$') share the same downstream deletion is remarkable. This shared polymorphism could be the result of an intragenic recombination event, gene conversion or, perhaps more likely, the result of the same slippage event occurring twice in the downstream region (Dover, 1989).

3.3.6 Sequence analysis of the perfect Thr-Gly encoding repeats of natural Australian populations.

Figure 3.5 shows the DNA sequences encoding the perfect repeat. The sequences are aligned according to the length of the uninterrupted repeat and the arrows are used to highlight different Thr-Gly encoding cassettes. At present there are 8 known length classes corresponding to 14, 17, 18, 20, 21, 22, 23 and 24 Thr-Gly pairs. Apart from the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{24}$ length variants all the other length classes have isolength alleles which have silent substitutions in the third positions of the Thr-Gly encoding ACN GGN hexamers. As reported in Costa *et al.*, 1992 and chapter 4 the $(Thr\text{-}Gly)_{17}$ and the $(Thr\text{-}Gly)_{20}$ alleles are the two most frequent length variants comprising 90% of natural European populations and 82% of Eastern Australian populations. The $(Thr\text{-}Gly)_{23}$ allele is the next most frequent allele followed by the rarer $(Thr\text{-}Gly)_{14}$. Interestingly it has recently

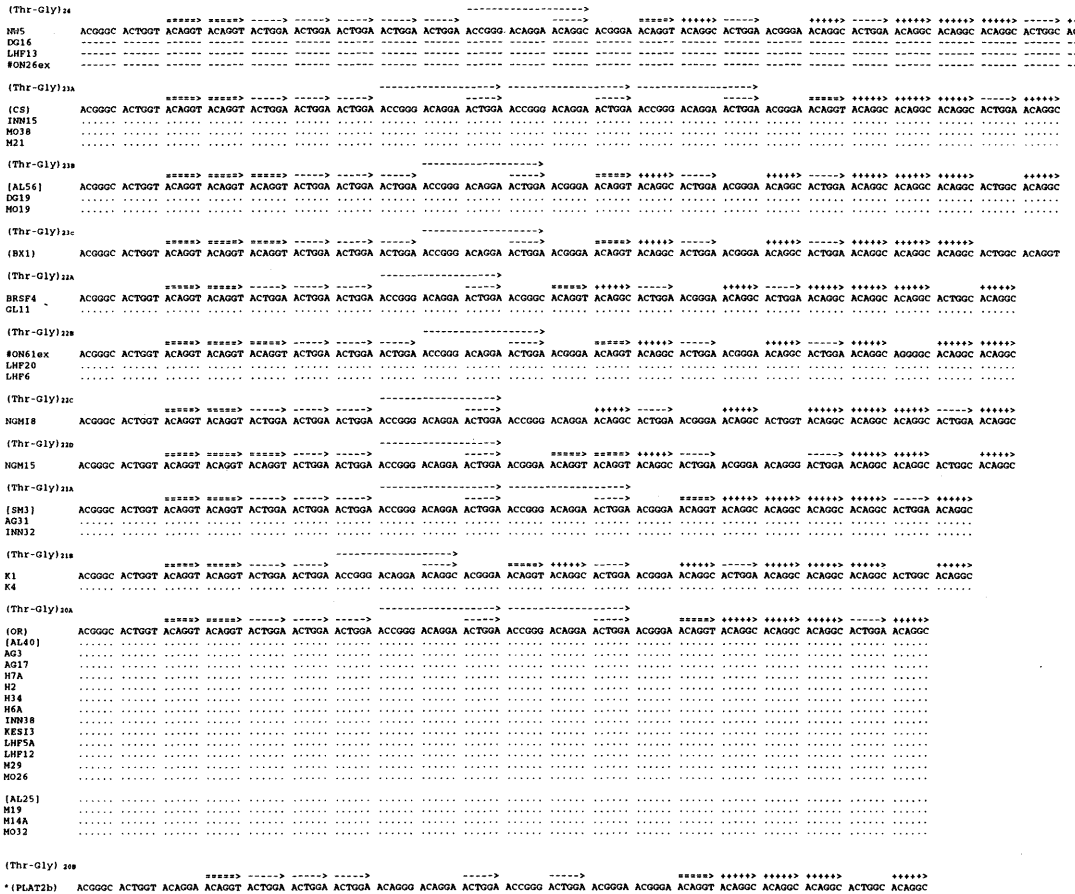


Figure 3.5. Nucleotide sequences of the perfect repeat of the *D. melanogaster* Thr-Gly region of natural Australian populations (see following page for full legend).

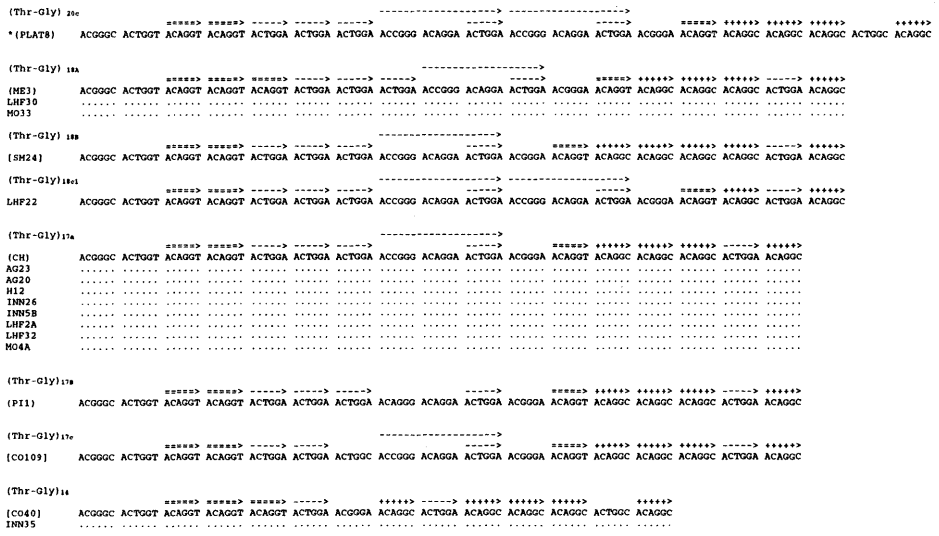


Figure 3.5. continued. Nucleotide sequences of the perfect repeat of the *D. melanogaster* Thr-Gly region of natural Australian populations. The sequences are ordered on the basis of the length of the region and according to the different alleles considered. Published sequences are reported in brackets ( ) (Costa *et al.*, 1991), those prefixed with * are from Rosato *et al*(1996) while the sequences described by Peixoto (PhD thesis 1993a) are shown in parentheses [ ]. Sequences prefixed with # are personal communications (C.Aquadro). The dots indicate conserved nucleotide positions. The arrows highlight similarities among the alleles (see text). AG = Yeppoon (Queensland) ,AL = Albaida (Spain), BRFS = Bowen (Queensland) , BX = Bordeaux (France), CO = Cognac (France), CON = Conselve (Italy), CS = Canton S, DG = S. Brisbane, (New South Wales), GL = Hervey Bay (Queensland) , H = Melbourne (Victoria), INN = Cairns (Queensland) , K = Kiambu, Nairobi (Kenya) , KESI = Kericho (Kenya), LHF = Yeppoon (Queensland) , M =Melbourne (Victoria) , ME = Merano (Italy), MO = Cairns (Queensland), NGMI = Nguruman (Kenya) , PI = Pietrastornina (Italy), PLAT = Platanistassa (Cyprus), SM = San Mateo (Spain).

been observed (E. Rosato and C. Pasetto pers. comm.) that the African $(Thr\text{-}Gly)_{23}$ length variant is as frequent as the European $(Thr\text{-}Gly)_{17}$ and the $(Thr\text{-}Gly)_{20}$ alleles.

At present there are three length alleles known for the $Thr\text{-}Gly_{17}$ variant class . The $(Thr\text{-}Gly)_{17a}$ allele which was first reported by Yu *et al.*, (1987b) and corresponds to the CH sequence, appears to be the most frequent and was found to be the only $(Thr\text{-}Gly)_{17}$ allele present in the lines analysed for behaviour (chapter 5, 6 and 7), apart from PI1 which has a $(Thr\text{-}Gly)_{17b}$ allele, differing from the $(Thr\text{-}Gly)_{17a}$ variant by a synonymous substitution (C $\rightarrow$ A, Fig. 3.5). DNA sequences of all eight alleles of the $(Thr\text{-}Gly)_{17a}$ variant (INN26, INN5b, MO4a, from Cairns; H12, from Melbourne; LHF2a, LHF32, AG23 and AG20, from Yeppoon) matched those found previously in the laboratory stock Chieti (CH) Yu *et al.*, 1987a). The third length allele.$(Thr\text{-}Gly)_{17c}$, was previously found in a Cognac population (CO109, Peixoto, PhD thesis 1993a).

There are also three different $(Thr\text{-}Gly)_{20}$ length alleles, two of which differ by a single substitution, $(Thr\text{-}Gly)_{20a}$ and $(Thr\text{-}Gly)_{20c}$ variants (A $\rightarrow$ C, Fig. 3.5). However $(Thr\text{-}Gly)_{20a}$ and $(Thr\text{-}Gly)_{20b}$ alleles differ by five transversions, so obviously some mechanism other than point mutation has been responsible for these changes. The $(Thr\text{-}Gly)_{20a}$ sequence which corresponds to the OR sequence (Citri *et al.*, 1987) is the most common and again was the only allele to be found within the lines analysed for locomotor activity (chapters 5,6 and 7). Fifteen $(Thr\text{-}Gly)_{20a}$ allele sequences (AG3, AG17, LHF5A, LHF 12, from Yeppoon; H7A, H34, H2, H6A, M14A, M19, M29 ,from Melbourne; KESI3 ,from Kericho and INN38, MO26, MO32, from Cairns) corresponding to the $(Thr\text{-}Gly)_{20a}$ Oregon-R (OR) sequence, were found (Citri *et al.*, 1987). Two $(Thr\text{-}Gly)_{20b/c}$ alleles have only ever been found in the Platanistassa sample from Cyprus (Rosato *et al.*, 1996).

The $(Thr\text{-}Gly)_{23}$ allele class again consists of three isolength alleles, and again two, $(Thr\text{-}Gly)_{23b}$ and $(Thr\text{-}Gly)_{23c}$, differ by only one base substitution (C $\rightarrow$ T, Fig. 3.5, BX1 previously described in Costa *et al.*, 1991). The sequences of five $(Thr\text{-}Gly)_{23}$ alleles from different populations (DG19, S. Brisbane; INN15,MO19, MO38, Cairns; M21, Melbourne) were also obtained. Three of these $(Thr\text{-}Gly)_{23}$ alleles shared the same sequences in the encoding repeat as for Canton-S, $(Thr\text{-}Gly)_{23a}$, INN15, MO38 and M21 (Jackson *et al.*, 1986; Yu *et al.*, 1987b; Wheeler *et al.*, 1991). Two were identical to AL56, $(Thr\text{-}Gly)_{23b}$

variant (Peixoto PhD thesis, 1993a), DG19 and MO19. However no $(Thr\text{-}Gly)_{23c}$ length allele identical to BX1 was observed (Costa *et al.*, 1991).

Two different isolength $(Thr\text{-}Gly)_{18}$ variants were previously found in Europe. One from Merano (ME3) denoted the $(Thr\text{-}Gly)_{18a}$ variant (Costa *et al.*, 1991), while the other from San Mateo (SM24) is denoted as the $(Thr\text{-}Gly)_{18b}$ allele (Peixoto PhD thesis 1993a). A third new $(Thr\text{-}Gly)_{18}$ isolength variant was found, LHF22, (Yeppoon, Queensland) known as $(Thr\text{-}Gly)_{18c}$ which differed from the $(Thr\text{-}Gly)_{18b}$ allele by eight transversions indicating mutational events other than point mutations occurring. The two $(Thr\text{-}Gly)_{14}$ alleles, CO40 from Cognac and INN35 from Cairns, do not show any sequence differences. A similar situation exists in the case of the three $(Thr\text{-}Gly)_{21}$ alleles, SM3 from Spain, AG31 from Yeppoon and INN32 from Cairns. However, one new isolength $(Thr\text{-}Gly)_{21}$ allele was observed when K1 and K4 from Nairobi were sequenced and denoted the $(Thr\text{-}Gly)_{21b}$, the first identified allele is now the $(Thr\text{-}Gly)_{21a}$ allele. The $(Thr\text{-}Gly)_{21b}$ allele differed from $(Thr\text{-}Gly)_{21a}$ by 13 base substitutions, again suggesting that it was derived from a different lineage. Four different isolength alleles of the rare variant $(Thr\text{-}Gly)_{22}$ were found, one the $(Thr\text{-}Gly)_{22b}$ allele had been observed before, ON61ex (Aquadro, unpublished observations) and two flies from Yeppoon (LHF20 and LHF6) also shared the same DNA sequence. The new $(Thr\text{-}Gly)_{22a}$ allele was observed when flies from Bowen (BRSF4) and Hervey Bay, (GL11) were sequenced. This new allele differed from the $(Thr\text{-}Gly)_{22b}$ variant by 21 nucleotide substitutions. The third and fourth $(Thr\text{-}Gly)_{22}$ isolength alleles were observed in single flies both emanating from Nguruman (Kenya), NGMI8, the new $(Thr\text{-}Gly)_{22c}$ and NGMI5 the novel $(Thr\text{-}Gly)_{22d}$ length variants respectively.

The arrows in Figs. 3.3 and 3.5, show putative deletion/duplication events can be identified that can be used to explain the origin of these different length variants. For example, Yu *et al.*, (1987b) have shown that the $(Thr\text{-}Gly)_{17a}$, $(Thr\text{-}Gly)_{20a}$ and $(Thr\text{-}Gly)_{23a}$ variants from laboratory stocks (CH, OR and CS, respectively) differ from each other by the presence of one, two or three copies respectively of an 18 base-pair perfect direct repeat (indicated by the arrow ----------> in Figs. 3.3 and 3.5) and consequently they can be derived from each other by simple deletion/duplication events. It can also be seen from Fig. 3.3 and 3.5 that the $(Thr\text{-}Gly)_{18a}$ variant found in the Merano population (ME3) and LHF30

and MO33 from Yeppon and Cairns respectively, differs from the $(Thr\text{-}Gly)_{17}$ variant by just a 6 base-pair duplication (ACAGGT, indicated by the arrow =====> in Figs. 3.3 and 3.5). This repeat appears in three tandem copies in ME3, LHF30 and MO33, and only in two in the $(Thr\text{-}Gly)_{17}$ alleles. The other $(Thr\text{-}Gly)_{18}$ length alleles can also be explained in a similar way, whereas $(Thr\text{-}Gly)_{18b}$ variant differs from $(Thr\text{-}Gly)_{17a}$ allele by a different 6bp duplication (ACAGGC, indicated by the arrow +++++> in Figs 3.3 and 3.5). However, not only do these rare length alleles appear to arise from the $(Thr\text{-}Gly)_{17}$ variant but the $(Thr\text{-}Gly)_{18c}$ length allele differs from the $(Thr\text{-}Gly)_{20a}$ variant by a deletion of 12bp (ACAGGC ACAGGC, indicated by +++++> +++++> in Fig. 3.5)

Fig. 3.6 shows the alignment of these different ACN GGN cassettes or hexamers as a schematic representation of the Thr-Gly allele sequences of Fig. 3.5. Each Thr-Gly pair is represented by a lower case letter, and numerical subscripts are used to indicate likely base substitutions which occur in repetitive units and which appear to share a common descent. For example 'd_1'(ACT GGC) is most likely to be a synonymous substitution in the 'd' cassette (ACT GGA). Costa *et al.* (1991) and Peixoto *et al.* (1992) suggested that the main perfect repeat of the *Drosophila* Thr-Gly region has largely evolved by the action of deletion/duplication events as commonly observed in repetitive sequences (Dover 1987, 1989; Levinson and Gutman 1987). In Fig. 3.6 , the sequences are aligned according to this assumption and I have attempted to minimise the number of synonymous substitutions between any two different length variants by assuming that only one deletion/duplication event occurred between any pair. It is important to note that in the majority of cases the deletion/duplication events could occur in both directions. Taking the previously mentioned example of the $(Thr\text{-}Gly)_{17a}$, $(Thr\text{-}Gly)_{20a}$, and $(Thr\text{-}Gly)_{23a}$ alleles, the deletion and duplication of the 'efd' element could occur in any direction, as with the $(Thr\text{-}Gly)_{18b}$ and $(Thr\text{-}Gly)_{21a}$ length alleles. Due to the large number of length alleles shown here it is almost impossible to postulate direct derivations without the inclusion of multiple events such as gene conversion, deletion/duplication and base substitutions. Therefore it was decided to arbitrarily group the alleles into different 'types' of repeat. It was previously suggested through flanking haplotype analysis of the different Thr-Gly length alleles (Rosato *et al.*, 1996) that the $(Thr\text{-}Gly)_{20a}$ and $(Thr\text{-}Gly)_{23b}$ are the ancestral alleles of this region.

Moreover, all the other alleles presented here could have arisen., from these two alleles by unequal crossover, by slippage like or unequal sister chromatid exchange events and by gene conversion-like mechanisms,

3.3.7 The possible derivations of the different Thr-Gly length variants.

When grouping all the known Thr-Gly length alleles into two discrete classes, whether '$(Thr\text{-}Gly)_{20a}$' like or '$(Thr\text{-}Gly)_{23b}$' like, the following table is produced, Table 3.1.

Table 3.1 Distribution of possible ancestral origins for each Thr-Gly length allele

original allele	'$(Thr\text{-}Gly)_{20a}$' like allele	'$(Thr\text{-}Gly)_{23b}$' like allele
$(Thr\text{-}Gly)_{24}$	✗	✓
$(Thr\text{-}Gly)_{23a}$	✓	✗
$(Thr\text{-}Gly)_{23c}$	✗	✓
$(Thr\text{-}Gly)_{22a}$	✗	✓
$(Thr\text{-}Gly)_{22b}$	✗	✓
$(Thr\text{-}Gly)_{22c}$	✓	✓
$(Thr\text{-}Gly)_{22d}$	✓	✓
$(Thr\text{-}Gly)_{21a}$	✓	✗
$(Thr\text{-}Gly)_{21b}$	✗	✓
$(Thr\text{-}Gly)_{20b}$	✓	✗
$(Thr\text{-}Gly)_{20c}$	✓	✗
$(Thr\text{-}Gly)_{(20a\text{-}2)}$	✓	✗
$(Thr\text{-}Gly)_{18a}$	✓	✗
$(Thr\text{-}Gly)_{18b}$	✓	✗
$(Thr\text{-}Gly)_{18c}$	✓	✗
$(Thr\text{-}Gly)_{17a}$	✓	✗
$(Thr\text{-}Gly)_{17b}$	✓	✗
$(Thr\text{-}Gly)_{(17c\text{-}2)}$	✓	✗
$(Thr\text{-}Gly)_{14}$	✗	✓

Generally the '$(Thr\text{-}Gly)_{20a}$' type alleles have either a complete duplication of the 'efd' cassette or the absence of the 'hdghd' cassette (Fig. 3.6), whereas the '$(Thr\text{-}Gly)_{23b}$' type alleles lack the duplicated 'efd' cassette and have a downstream 'hdghd' cassette. Those alleles that appear to be a mixture of both are the African flies, where intra-allelic exchange may have occurred between the two different length alleles. It is interesting to observe these mix/match alleles have only been found in Africa, presumably the ancestral population,

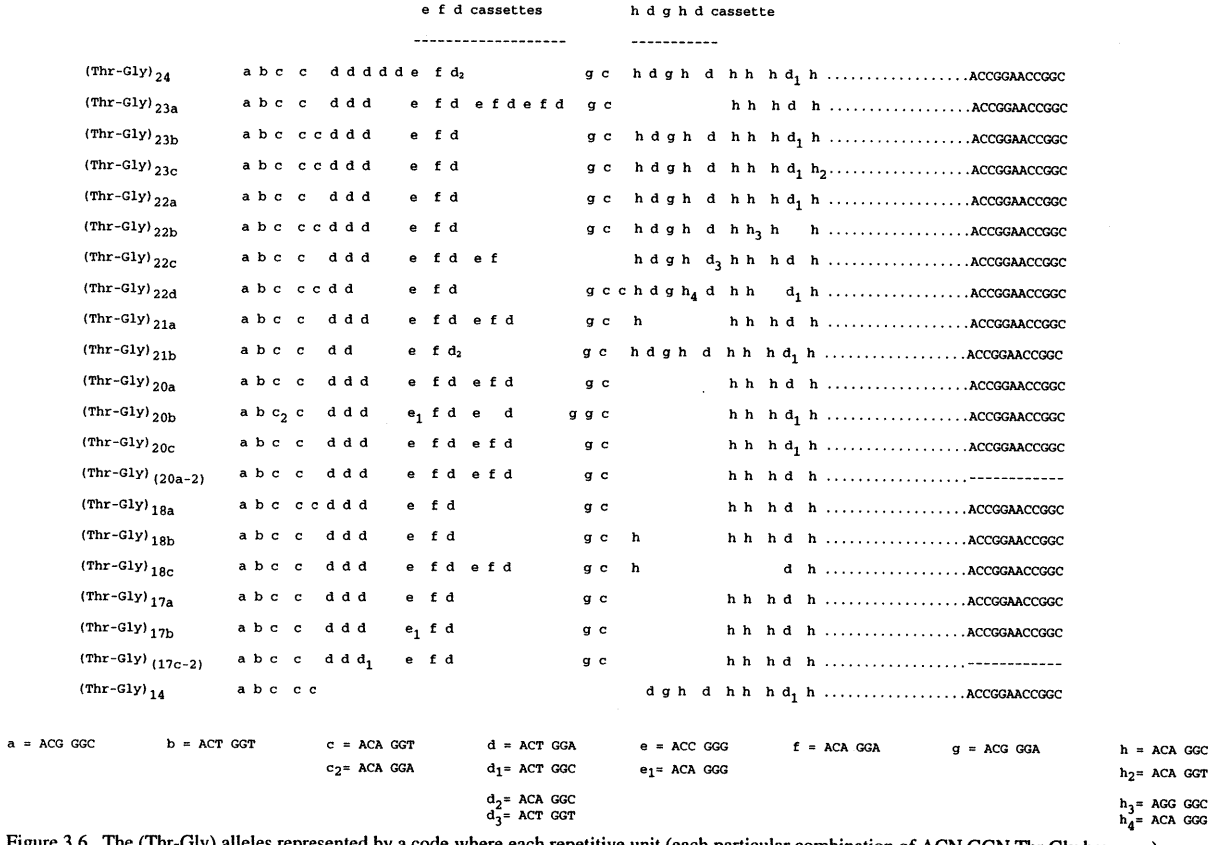


Figure 3.6. The (Thr-Gly) alleles represented by a code where each repetitive unit (each particular combination of ACN GGN Thr Gly hexamer) corresponds to a lower case letter. Subscripts (for example "d_1" etc.) denote Thr-Gly encoding "cassettes" which appear to share common descent. The alignment of the sequences is based on the hypothesis that deletion/duplication and base substitution events are responsible for the evolution of the (Thr-Gly) main repeat. The downstream $(Thr\text{-}Gly)_2$ is also included, dashes indicate deletions.

where greater allelic variation exists. Since several of the alleles presented here are exact replicates of those from Rosato *et al.* (1996) the postulated evolution of the Thr-Gly alleles reported by these authors will be omitted here. Only the new length alleles found in Europe, Africa and Australia will be discussed.

For example, using Fig. 3.6 as a guide, the $(Thr\text{-}Gly)_{24}$ allele can be plausibly derived from the $(Thr\text{-}Gly)_{23b}$ allele by the deletion of a 'c' unit, a duplication of a 'dd' unit and a base substitution of 'd' to 'd_2'. One of the new $(Thr\text{-}Gly)_{22}$ length variants found in Australia (LHF $(Thr\text{-}Gly)_{20}$ and LHF 6), the $(Thr\text{-}Gly)_{22b}$ variant, can be easily seen to have originated by two deletions ('c' and 'd_1' units) plus a base substitution ('h' to 'h_3'). The other non intragenic recombinant of this Thr-Gly length allele, the $(Thr\text{-}Gly)_{22a}$, only differs from the $(Thr\text{-}Gly)_{23b}$ allele by the single 'c' unit deletion. A second $(Thr\text{-}Gly)_{21}$ allele, the $(Thr\text{-}Gly)_{21b}$ variant, differs from the ancestral $(Thr\text{-}Gly)_{23b}$ allele by the same 'c' unit deletion, plus a 'd' unit deletion followed by a 'd' to 'd_2' base substitution. The single novel allele found which appears to originate from the other possible ancestral allele, the $(Thr\text{-}Gly)_{20a}$ variant, is the $(Thr\text{-}Gly)_{18c}$ allele, which differs from the former by an 'hh' unit deletion.

Synonymous substitutions between isolength alleles can also be accounted for by deletion and duplication events from different ancestral allele lineages. For example, the 13 base substitutions which are observed between the $(Thr\text{-}Gly)_{21a}$ and $(Thr\text{-}Gly)_{21b}$ alleles can be explained if we presume the $(Thr\text{-}Gly)_{21a}$ allele was derived from the $(Thr\text{-}Gly)_{20a}$ ancestral allele by a 'h' unit deletion, whereas, the $(Thr\text{-}Gly)_{21b}$ variant may have been derived from the $(Thr\text{-}Gly)_{23b}$ allele by a deletion of a 'c' cassette, a duplication of a 'd' unit and a two base substitutions from a 'd' to 'd_2' cassette (Fig. 3.6).

3.3 DISCUSSION

Repetitive regions are particularly interesting for studying molecular evolution. Repetitive DNA arrays are prone to mispairing and show relatively high mutation rates (Jeffreys *et al.*, 1988a, 1990) that are generated by several mutational mechanisms, such as

slippage and unequal crossover (Dover, 1987, Eichler *et al.*, 1995). Consequently they are usually polymorphic in length and can be used as DNA markers, such as the noncoding mini- and microsatellites (Jeffreys *et al.*, 1993; Tautz *et al.*, 1989), and for studying the mechanisms for generating new allelic variants (Jeffreys *et al.*, 1993; Strand *et al.*, 1993). Entire genes can also be repeated and tandem arrays of repeats can also be found within genes. It has recently been demonstrated that the expansion of trinucleotide repeats encoding for a polyglutamine array can be the cause of several human neurodegenerative disorders (Richards and Sutherland; 1992; Kuhl and Caskey, 1993; Housman, 1995). Internally repetitive genes often show intraspecific length polymorphism in coding DNA as, for example, in the *Drosophila dec-1* (Andersson and Lambertsson, 1993), and *notch* genes (Tautz, 1989), and in the gorilla and human involucrin genes (Green and Djian, 1992). For other genes, like *per*, the repetitive region consists of a dipeptide (Jackson *et al.*, 1986; Yu *et al.*, 1987b; Citri *et al.*, 1987; Costa *et al.*, 1991; Peixoto *et al.*, 1992).

The existing phylogenetic model of Thr-Gly repeat evolution (Peixoto PhD thesis, 1993a, Rosato *et al.*, 1996) predicts that the $(Thr\text{-}Gly)_{20a}$ and $(Thr\text{-}Gly)_{23b}$ variants are representative of the 'ancestral' state compared to the others, because they can all be derived from these two variants under the assumptions made. Obviously there are other possible scenarios for more complicated patterns if different putative intermediate variants are hypothesised. Gene conversion and unequal crossing-over between different length variants might play a significant role in the evolution of this region. Indeed if these processes occur frequently then the relationship between the different variants could be quite different. Evidence for gene conversion was found when examining the pattern of base substitutions amongst the alleles (Rosato *et al.*, 1996), where regularity can be highlighted (Fig. 3.6). For example a 'd' to 'd_1' substitution is the most common one, occurring as the penultimate unit in several length alleles. The 'e' to 'e_1' substitution occurs in both the $(Thr\text{-}Gly)_{17b}$ and $(Thr\text{-}Gly)_{20b}$ variants while the 'c' to 'c_2', 'd' to 'd_3', 'h' to 'h_2', 'h_3' and 'h_4' are all unique. The 'd' to 'd_2' mutation occurs in two newly found length variants, the $(Thr\text{-}Gly)_{24}$ and $(Thr\text{-}Gly)_{21b}$. The sharing of mutations between repetitive units belonging to either different alleles or to the same one is a frequent phenomenon, for example, in the mammalian involucrin gene (Green and Djian 1992), in the late chorion locus of *Bombyx mori*

(Eickbush and Burke 1986) and in the *dec-1* eggshell locus in *Drosophila* (Andersson and Lambertsson 1993). This spread of mutations is known as 'correction', and has been reported for isolength and different length involucrin alleles (Green and Djian 1992, Djian *et al.*, 1993). Correction without length variation implies either a double recombination or a more parsimonious gene conversion. This latter mutational event could explain the origin of the 'd_1' and 'e_1', as well as many other possible new repeat units.

Rosato *et al.*(1996) also constructed an alternative genealogy of the Thr-Gly repeats using the proximal flanking markers at positions 5101, 5242 and 5247, where four combinations of bases occur at these sites, GGC, GAC. GAT and CAC, respectively. Substitutions at these sites are less likely to occur than length changes within the repeat, therefore the tree was constructed through clustering of these haplotypes in such a way as to minimise changes at these sites. The tree differed to the one previously constructed based on the ancestral alleles of $(Thr\text{-}Gly)_{20a}$ and $(Thr\text{-}Gly)_{23b}$. The GGC haplotype was taken as ancestral since these bases also occur in *D. simulans* , however, using this to root the tree meant the $(Thr\text{-}Gly)_{23b}$ was no longer considered to be ancestral since it was in complete linkage disequilibrium with GAT, implying two substitution events from the GGC ancestral state. However, this perfect linkage is no longer true since the observation of the $(Thr\text{-}Gly)_{23b}$ line (DG9) having a GAC flanking haplotype (Fig. 3.4), is only one base substitution away from the ancestral haplotype. Because other *Drosophila* species show the 'hdghd' $(Thr\text{-}Gly)_{23b}$ element (Peixoto *et al.*, 1992; Rosato *et al.*, 1994), this alternative genealogy seems to suggest some form of evolutionary convergence between species.

To observe a high level of variability in natural populations, the existence of mechanisms creating polymorphism is not enough. The new variants need to be at least functionally equivalent or they are rapidly eliminated from the populations by natural selection. Repetitive coding arrays therefore differ from noncoding repetitive DNA, which seems to be maintained only by its ability to replicate with the rest of the genome until its excessive expansion becomes visible to selection (Charlesworth *et al.*, 1994).

By detailed examination of the Thr-Gly sequences we were able to identify the action of several DNA turnover mechanisms acting on the Thr-Gly main repeat, in particular, with respect to previous studies which emphasised the action of slippage-like and unequal cross-over mechanisms (Costa *et al.*, 1991, Peixoto *et al.*, 1992, Peixoto, PhD thesis 1993a, Rosato *et al.*, 1996). It was also observed that the majority of variants have a perfect Thr-Gly repeat followed by a $(Thr\text{-}Gly)_2$ repeat upstream, and that the deletion of

this upstream motif is very rare. The model that states the $(Thr\text{-}Gly)_{20a}$ and $(Thr\text{-}Gly)_{23b}$ are the ancestral alleles is supported by this sequence analysis presented here since new alleles found can be derived from one or other of these.

CHAPTER 4

THE EUROPEAN CLINE REVISITED AND SPATIAL ANALYSIS OF THE Thr-Gly LENGTH POLYMORPHISM IN AUSTRALIAN POPULATIONS OF *D. melanogaster*.

4.1 INTRODUCTION

The Thr-Gly region is polymorphic in length in both laboratory and natural populations (Yu *et al.*, 1987b; Costa *et al.*, 1991). Furthermore the two major length variants in Europe, the $(Thr\text{-}Gly)_{17}$ and the $(Thr\text{-}Gly)_{20}$ show a highly significant geographical differentiation, producing a robust latitudinal cline, the former allele predominant in the south and the latter in the north of Europe (Costa *et al.*, 1992). Traditionally, a latitudinal cline may suggest the action of natural selection, with some of the strongest evidence for selectively driven clines coming from the observations of similar latitudinal clines for allelic variation in alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase on different continents (Oakeshott *et al.*, 1982). If the European cline is being maintained by selection then historical factors that could possibly explain this spatial patterning need to be eliminated.

At least three evolutionary processes, or combinations thereof, may determine allele frequency gradients on a continental scale. The first, demic diffusion, has been observed in humans (Menozzi *et al.*, 1978) where population growth, centrifugal dispersion and admixture in an area of expansion leads to stable clines. For European *Drosophila* there is no evidence that one or few local populations have undergone dramatic increase in numbers leading to an expansion into new regions. The second explanation that could possibly account for the observed distribution in Europe is range expansion by a founder allele. The random genetic differences thus established may lead to clines in successive gene flow. This process has been proposed for polymorphic loci in the giant toad *Bufo marinus* (Easteal, 1985). The most conventional explanation for a latitudinal cline is the adaptive response to climatic variation (Endler, 1977: Piazza *et al.*, 1981; Sokal *et al.*, 1987). Latitude selection pressures are thought to be responsible for the clines at the *alcohol dehydrogenase* (*Adh*) and *esterase-6* (*Est6*) loci in *D. melanogaster* (Oakeshott *et al.*, 1982: Anderson and Oakeshott 1984, David and Capy,

1988; David *et al.*, 1989) as well as in the chromosomal inversion polymorphism of *D. subobscura* (Ayala *et al.*, 1989). It was first thought that the climatic association of maximum temperature was responsible for the North American clines of *Adh* and *est6* (Anderson, 1981). However this hypothesis was rejected when examining flies from Australasia and Asia (Oakeshott *et al.*, 1982), whose populations exhibited clines with latitude but no correlations were found with temperature to support a relationship with temperature gradients. Geographic variation in *Adh* allele frequencies of all three continents is related to maximum rainfall, but the mechanisms underlying this relationship are difficult to explain (Oakeshott *et al.*, 1982).

If selection is directly acting on the Thr-Gly repeat, temperature could be the selective agent. However, no significant correlation of the Thr-Gly allele frequencies was found with European temperature data for the cline (Costa *et al.*, 1992). Deletion of the Thr-Gly repeat plus some flanking amino acids, resulted in a lengthening of the circadian period at higher temperatures in per^{01} transformants (Ewer *et al.*, 1990). This result might indicate that the Thr-Gly region could be involved in the temperature compensation of the circadian phenotype(see chapter 5). However it is not surprising that a large deletion within a protein may make it more sensitive to temperatures via non specific conformational changes, so the significance of this result is unclear.

In this chapter a further analysis of the European Thr-Gly allele frequencies with respect to temperature is undertaken, using a more extensive data set of environmental factors for the European continent (Hulme *et al.*, 1995, Climate Impacts LINK Project). Secondly a significant correlation between the Thr-Gly length polymorphism and latitude and/or temperature involving a different continent would provide strong evidence for the action of selection. If the European and North African cline was simply caused by historical processes unrelated to selection, then it is unlikely to be duplicated in the southern hemisphere. A geographical analysis of the Thr-Gly repeat length polymorphism in natural populations collected from a 2600km north - south transect in eastern Australia (James and Partridge., 1995) is reported.

4.2 METHODS

Single fly DNA preparation, PCR and gel electrophoresis were carried out as described in chapter 2. Male flies taken from isofemale lines were immediately frozen and analysed. Samples from 20 natural populations of *D. melanogaster* were collected at sites on a 2600km transect along the eastern coast of Australia during February 1993 and kindly donated by Avis James and Linda Partridge. The collections were sampled from 13 latitudes, seven of these with two replicate sites (Fig. 4.1, Table 4.5). The nearest weather stations were noted and used for both latitude and longitudinal coordinates of the fly populations and the relevant recorded temperature data was referenced (Gentilli, 1971). The climatology data for Europe was made available by the Climate Impacts LINK Project (Hulme *et al.*, 1995).

The computations for spatial autocorrelation were carried out using the SAAP program developed by David Wartenberg (Version 4.3, October 1989) and the correlations by Statistica (Statsoft).

4.3 RESULTS

A. Revisiting the thermal correlations for the European cline

It was surprising that Costa *et al.* (1992) found no significant correlation for any of the six climatic parameters, particularly temperature, with the Thr-Gly allele frequencies of the 18 localities examined, since correlations have been observed between temperature and latitude at least in Australia (James and Partridge, 1995). The correlation between the mean yearly temperatures for the European localities examined here and their latitude is also highly significant ($r = -0.84$, $p<<0.01$). However, the temperature data used by Costa *et al.* (Espenshade and Morrison, 1986) was not up-to-date and estimated very approximately from

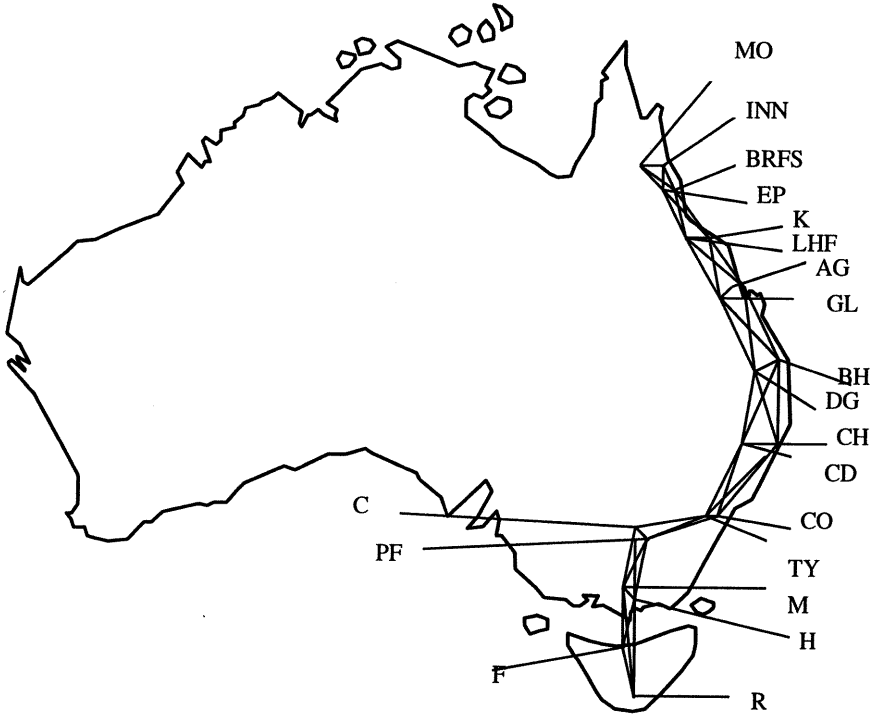


FIGURE 4.1 Map of the Eastern coast of Australia and the sampled populations. Codes were assigned as in Table 4.5. Localities are connected by a Delaunay network, to allow calculation of the Royaltey-Astrachan-Sokal statistic. Positions of populations are slightly exaggerated to show the network .

the nearest weather station of each locality. A more extensive temperature data set defined by latitudes 25° N to 81° N and by longitudes 32° W to 66° E, constructed from observed weather station data distributed across the region and interpolated to a resolution of 0.5° latitude/longitude land cells, resulting in a gridlike network of 12500 cells, was used in this analysis. The accuracy of the various interpolations was assessed using independent station data not used in the interpolation and this was shown to be between 0.5°C and 0.8°C for surface temperatures (Hulme *et al.*, 1995).

Initially each population was transformed into 'mean Thr-Gly repeat length', so for example Casablanca with 76.2% $(Thr\text{-}Gly)_{17}$, 4.7% $(Thr\text{-}Gly)_{18}$, 9.6% $(Thr\text{-}Gly)_{20}$ and 9.6% $(Thr\text{-}Gly)_{23}$ gave a 'mean' of 17.93. This was then correlated with the mean yearly temperature for each locality. Even though Thr-Gly repeat length is a discontinuous variable the majority of the alleles are either $(Thr\text{-}Gly)_{17}$ or $(Thr\text{-}Gly)_{20}$, therefore the 'mean Thr-Gly repeat length' describes reasonably accurately the relative $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ frequencies. A significant correlation coefficient was observed, $r = -0.544$, $p = 0.034$, (Fig. 4.2), so as the temperature increases the frequency of the shorter Thr-Gly length repeats increases. Correlations were also performed with each of the three most common alleles against mean yearly temperatures (Table 4.1), however the only significant correlation was seen with the $(Thr\text{-}Gly)_{17}$ variant.

Table 4.1 Pearson product moment correlations of the Thr-Gly alleles frequencies with respect to mean yearly temperatures °C (*<0.05)

Allele	*r*	p value
$(Thr\text{-}Gly)_{17}$	0.51	0.003*
$(Thr\text{-}Gly)_{20}$	-0.37	0.136
$(Thr\text{-}Gly)_{23}$	-0.38	0.120

When the mean yearly temperatures were broken down into monthly variables and correlated with the mean Thr-Gly repeat length across the European localities, significant relationships

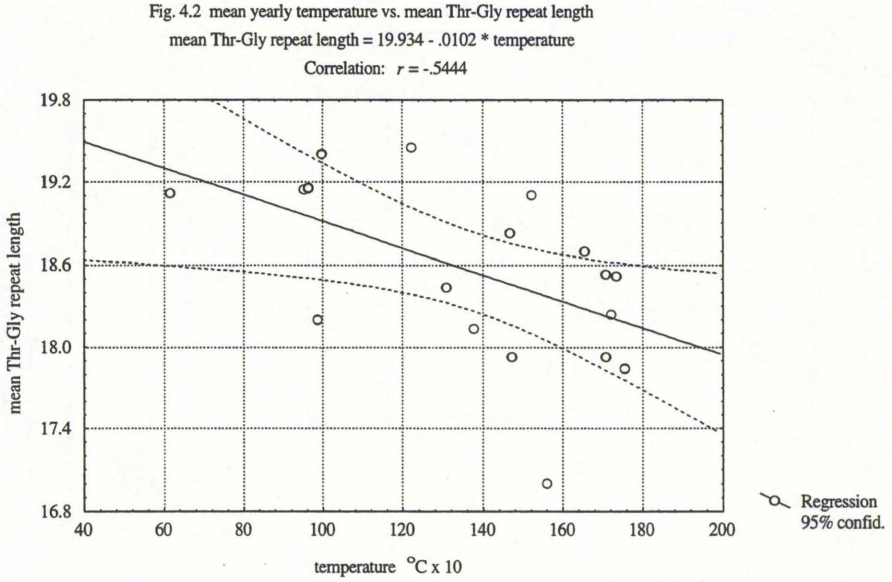


Fig. 4.2 Correlation of mean yearly temperature versus mean Thr-Gly repeat length for European populations

Table 4.2 Pearson product moment correlations of the 'mean Thr-Gly repeat length' with respect to mean monthly temperatures ^{o}C (*<0.05)

Month	*r*	p value
January	-0.54	0.021*
February	-0.54	0.021*
March	-0.51	0.030*
April	-0.48	0.044*
May	-0.46	0.054
June	-0.45	0.062
July	-0.44	0.067
August	-0.50	0.037*
September	-0.55	0.019*
October	-0.56	0.017*
November	-0.60	0.018*
December	-0.54	0.019*

were observed for all months except May, June and July, which narrowly missed significance (Table 4.2). It was interesting to observe that each correlation coefficient was negative, again suggesting an increase in mean Thr-Gly with a decrease in temperature across localities.

The frequencies of each common Thr-Gly length variant were also correlated with the mean monthly temperatures. Firstly the $(Thr\text{-}Gly)_{17}$ was significantly positively correlated with the late spring and summer months, April to November (Table 4.3). In contrast the $(Thr\text{-}Gly)_{20}$ frequencies showed no significant correlation with the mean monthly temperatures, and the $(Thr\text{-}Gly)_{23}$ frequencies were significantly negatively correlated for a few winter months, December, January and February. The observation of the positive correlations of the mean summer monthly temperatures with the frequency of the $(Thr\text{-}Gly)_{17}$ allele, which is found predominantly in southern Europe, does suggest that this length allele is particularly adapted to warmer climates.

Correlations were repeated with mean monthly sunshine hours and similar results were observed, which is not surprising as surface temperature is largely related to solar radiation.

Table 4.3 Pearson product moment correlations of the common Thr-Gly allele frequencies with respect to mean monthly temperatures ^{o}C (*<0.05)

	$(Thr\text{-}Gly)_{17}$		$(Thr\text{-}Gly)_{20}$		$(Thr\text{-}Gly)_{23}$	
Month	*r*	p value	*r*	p value	*r*	p value
January	0.44	0.070	-0.22	0.386	-0.54	0.020*
February	0.46	0.058	-0.28	0.264	-0.47	0.049*
March	0.46	0.053	-0.31	0.206	-0.4	0.101
April	0.51	0.030*	-0.42	0.081	-0.22	0.377
May	0.54	0.021*	-0.46	0.056	-0.15	0.553
June	0.51	0.030*	-0.45	0.062	-0.13	0.613
July	0.47	0.048*	-0.41	0.094	-0.16	0.531
August	0.50	0.034*	-0.41	0.091	-0.23	0.362
September	0.52	0.026*	-0.41	0.094	-0.32	0.199
October	0.51	0.030*	-0.34	0.161	-0.41	0.088
November	0.54	0.020*	-0.37	0.135	-0.45	0.059
December	0.45	0.059	-0.23	0.364	-0.54	0.020*

Longitudinal correlations of the Thr-Gly frequencies

The common Thr-Gly repeat length frequencies were also correlated across longitude (Table 4.4). No significant correlation was observed for the Thr-Gly allele frequencies, thus showing that the major directional component of the previously observed clinal patterns is north-south (Costa *et al.*, 1992).

Table 4.4 Pearson product moment correlations of the Thr-Gly alleles frequencies with respect longitude (*<0.05)

Allele	*r*	p value
$(Thr\text{-}Gly)_{17}$	0.39	0.113
$(Thr\text{-}Gly)_{20}$	-0.37	0.133
$(Thr\text{-}Gly)_{23}$	0.10	0.688

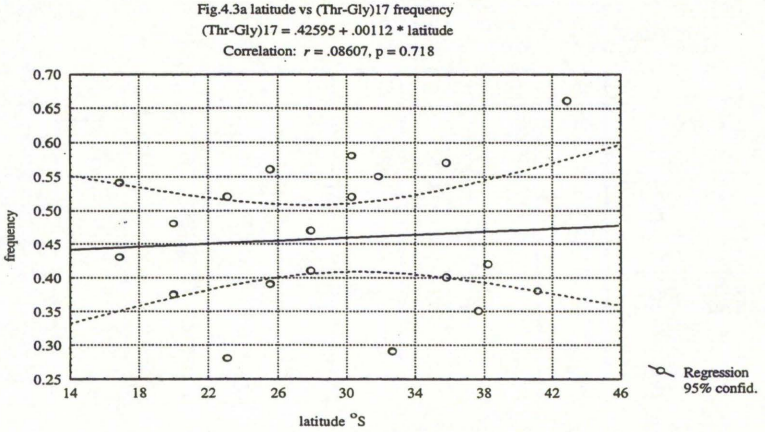
The results of these analyses suggests that temperature could be a selective agent in European populations for maintaining a clinal distribution of Thr-Gly length variants, especially since the frequency of the $(Thr\text{-}Gly)_{17}$ increases with an increase in temperature over the summer months. In contrast, the frequency of the $(Thr\text{-}Gly)_{20}$ allele increases with a decrease in temperature, although the correlations are not significant.

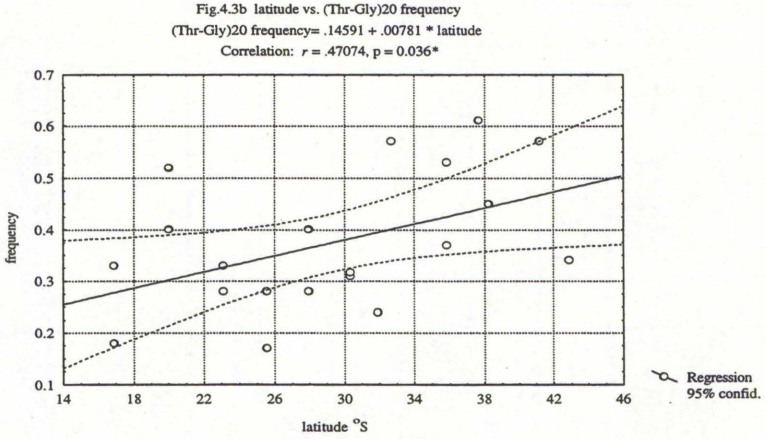
B. Latitudinal cline in the Thr-Gly length polymorphism of Australian *D. melanogaster*

Table 4.5 shows the frequencies of the different Thr-Gly length variants in 20 populations of Eastern Australia determined by PCR and agarose gel electrophoresis. Length variants corresponding in mobility to alleles with 14, 17, 18, 20, 21, 22, 23 and 24 Thr-Gly pairs were found. It is important to note here that, as mentioned in chapter 2, the primers used in the PCR amplified not only the uninterrupted Thr-Gly repeat but also the immediately downstream flanking sequence. These sequences also encode two downstream Thr-Gly pairs

POPULATION	CODE	DEGREES LATITUDE	DEGREES LONGITUDE	N	$(Thr\text{-}Gly)_{14}$	$(Thr\text{-}Gly)_{17}$	$(Thr\text{-}Gly)_{18}$	$(Thr\text{-}Gly)_{20}$	$(Thr\text{-}Gly)_{21}$	$(Thr\text{-}Gly)_{22}$	$(Thr\text{-}Gly)_{23}$	$(Thr\text{-}Gly)_{24}$
CAIRNS	MO	16.88	145.75	42	0.05	0.43	0.05	0.33	0.07	0	0.05	0.024
CAIRNS	INN	16.88	145.75	39	0.03	0.54	0	0.18	0.08	0	0.18	0
BOWEN	BRFS	20.02	148.25	32	0	0.38	0	0.4	0	0.03	0.19	0
BOWEN	EP	20.02	148.25	26	0	0.48	0	0.52	0	0	0.04	0
YEPPOON	K	23.13	150.75	27	0.07	0.52	0.04	0.33	0	0	0.04	0
YEPPOON	LHF	23.13	150.75	29	0.03	0.28	0.1	0.28	0.14	0	0.17	0
HERVEY BAY	AG	25.55	152.68	36	0	0.56	0.06	0.17	0.06	0	0.17	0
HERVEY BAY	GL	25.55	152.68	36	0	0.39	0	0.28	0	0.03	0.31	0
S.BRISBANE	BH	27.95	153.4	30	0	0.47	0.1	0.40	0	0.06	0.03	0
S.BRISBANE	DG	27.95	153.4	32	0	0.41	0.06	0.28	0	0	0.03	0.16
COFF'S HARBOUR	CH	30.32	153.12	29	0	0.52	0	0.31	0.04	0	0.10	0.04
COFF'S HARBOUR	CD	30.32	153.12	19	0	0.58	0.05	0.32	0	0	0.05	0
TAREE	CO	31.9	152.48	29	0.03	0.55	0.07	0.24	0	0	0.10	0
TAREE	TY	32.7	151.47	14	0	0.29	0.07	0.57	0	0	0.07	0
COBRAM	C	35.82	145.57	30	0.03	0.4	0.03	0.53	0	0	0	0
COBRAM	PF	35.82	145.57	35	0	0.57	0	0.37	0	0	0.06	0
MELBOURNE	M	37.68	145.53	23	0	0.35	0	0.61	0	0	0.04	0
MELBOURNE	H	38.23	145.03	33	0.09	0.42	0	0.45	0.03	0	0	0
TAZMANIA	F	41.18	146.37	21	0	0.38	0.05	0.57	0	0	0	0
TAZMANIA	R	42.88	147.33	38	0	0.66	0	0.34	0	0	0	0
TOTAL(N)	--	-------	--------	600	0.018(1)1	0.47(279)	0.032(19)	0.36(214)	(0.023)14	(0.007)4	(0.087)52	(0.012)7

TABLE. 4.5. Thr-Gly length variant frequencies in 20 natural Australian populations of *D. melanogaster*. N is the number of Thr-Gly alleles determined. Population codes were assigned as follows, High Falls Farm (MO), Innisfal Banana Farm (INN), Big Red Fruit Stand (BRFS), El Pedro Car Park (EP), Koppel Farm (K), Lazy Harry's Farm (LHF), Agnus Farm (AG), Goodlife Pools (GL), Brunswick Heads (BH), Dead Goose Farm (DG), Coff's Harbour (CH), Corindi (CD), Coopernook (CO), Tyrells (TY), Cornish Farm (C), Pullars Farm (PF), Chappies Farm (M), Hastings Farm (H), Forth (F) and Ranelagh (R). These populations are previously described in James and Partridge.,(1995), where INN is IN, BRFS is BS, K is KL, LHF is LH, CD is CI, C is CS, M is ME, H is HS, F is FT and R is RN.





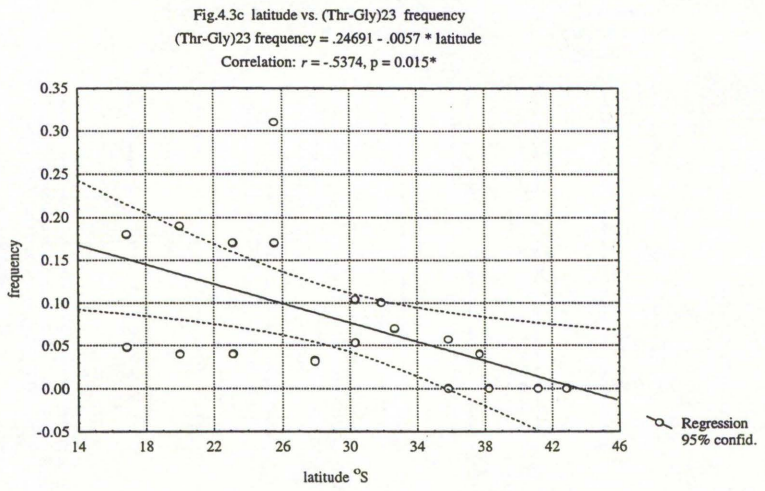


Figure 4.3 a, b and c Frequency plots of the $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ and variants with respect to latitude.

that were known to be deleted in one sequenced *D. simulans per* allele (Wheeler *et al.*, 1991). Therefore, in Table 4.5 the frequencies of Thr-Gly variants refer to the 'predicted' number of Thr-Gly pairs based on the length of the PCR fragment assuming that the two Thr-Gly pairs downstream of the uninterrupted repeat are not deleted.

The localities in Table 4.5 are ordered by latitude. As in the European and North Africa data the most frequent alleles are represented by the $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and the $(Thr\text{-}Gly)_{23}$ variants, making up approximately 47%, 36% and 9% of the Australian population respectively. As with the European population Thr-Gly length allele frequencies ($(Thr\text{-}Gly)_{17}$ 54%, $(Thr\text{-}Gly)_{20}$ 32%, Costa *et al.*, 1992), the $(Thr\text{-}Gly)_{17}$ is more common than the $(Thr\text{-}Gly)_{20}$ allele. This could be the result of selection favouring the $(Thr\text{-}Gly)_{17}$ allele generally across the continents, or particularly in Australia, the result of founder effects followed by migration. In addition at least one of the other rarer alleles, $(Thr\text{-}Gly)_{14}$ (2%) , $(Thr\text{-}Gly)_{18}$ (3%) , $(Thr\text{-}Gly)_{21}$ (2%) , $(Thr\text{-}Gly)_{22}$ (0.7%) and $(Thr\text{-}Gly)_{24}$ (1%) was also present in most of the populations, with the exception of the two originating from Tasmania. Inspection of the data in Table 4.5 suggests that populations at higher latitudes have higher frequencies of the $(Thr\text{-}Gly)_{20}$ variant than populations at lower latitudes. This is confirmed by the correlation between $(Thr\text{-}Gly)_{20}$ frequency and latitude (Table 4.6, Fig. 4.3b). However unlike Costa *et al.*, (1992) no significant complementary trend is observed for the $(Thr\text{-}Gly)_{17}$ allele (see Table 4.6, Fig. 4.3a). The $(Thr\text{-}Gly)_{23}$ variant was also significantly negatively correlated with latitude ($r = -0.54$, $p<0.05$, Table 4.6, Fig. 4.3c), whereas in the previous European study no significant correlation was observed with this length variant (Costa *et al.*, 1992). Table 4.6 gives all the correlations for the Australian Thr-Gly length variants. The $(Thr\text{-}Gly)_{21}$ frequencies were also significantly correlated with latitude, but as the majority of localities had zero frequencies for this length allele (Table 4.5), this result is probably meaningless.

Table 4.6 Pearson product moment correlations of the frequency of the Thr-Gly variants with respect to latitude (*, p<0.05)

Allele	*r*	p value
$(Thr\text{-}Gly)_{14}$	- 0.10	0.670
$(Thr\text{-}Gly)_{17}$	0.09	0.718
$(Thr\text{-}Gly)_{18}$	- 0.10	0.671
$(Thr\text{-}Gly)_{20}$	0.47	0.036*
$(Thr\text{-}Gly)_{21}$	- 0.45	0.046*
$(Thr\text{-}Gly)_{22}$	- 0.19	0.422
$(Thr\text{-}Gly)_{23}$	- 0.54	0.015*
$(Thr\text{-}Gly)_{24}$	- 0.08	0.724

To further dissect these initial correlations, the frequencies of the $(Thr\text{-}Gly)_{17}$, and the $(Thr\text{-}Gly)_{20}$ allele were compared with the cumulative frequencies of all other alleles across localities using a G-test (Sokal and Rohlf, 1981). This test examines whether allele frequencies are homogenously distributed, i.e. that the observed data represent random samples from a population with a single set of gene frequencies. The $(Thr\text{-}Gly)_{23}$ could not be included in this analysis as a separate allele but was pooled with the rest of the rarer variants as it was not present in all localities. A phenetic criterion was used, i.e. alleles were considered different if their Thr-Gly repeats differed in length (Excoffier *et al.*, 1992), so differences in DNA sequences at the synonymous position were not taken into consideration. The populations that had been collected at replicate latitudes were pooled (replicate populations from different latitudes from Taree and Melbourne were not pooled, see table 4.5) and the Tasmanian populations excluded because only one fly in the two Tasmanian populations represented an allele other than the two common classes (see table 4.5). This pooling and exclusion yielded 33 overall frequency classes which were then compared. The G-test was highly significant ($G = 46.18$, df, 10, $p < 0.001$), showing that the observed gene frequencies do not represent random samples drawn from a unique, panmictic or genetically uniform population.

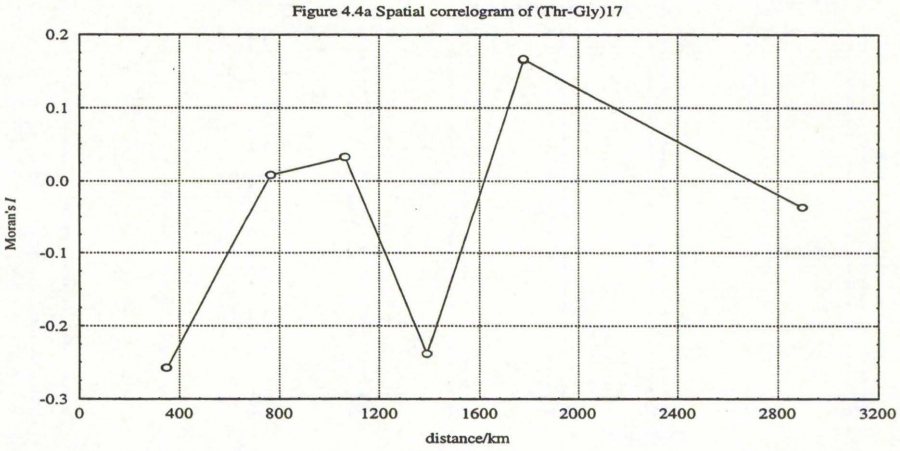
Spatial autocorrelation reflects the dependence of the value of a particular variable at one location on the value of that same variable at other nearby locations. Allele frequency

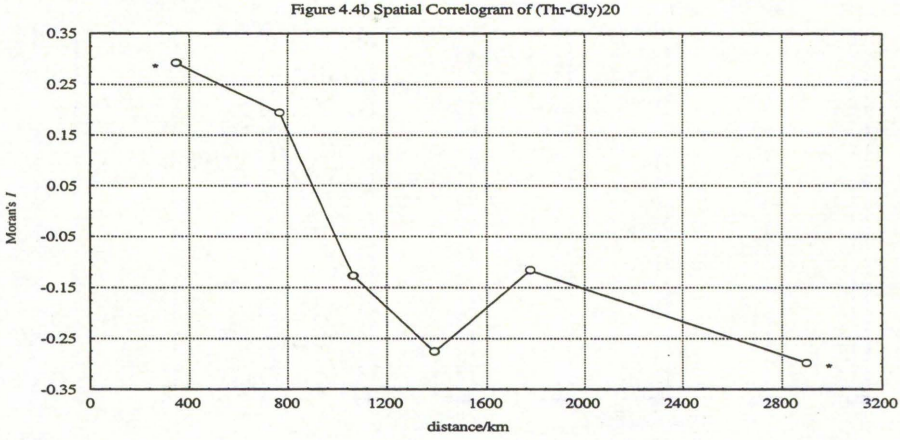
patterns were summarised by a spatial autocorrelation statistic, Moran's *I* (Sokal and Oden 1978) which represents the degree of similarity between populations as a function of their distance apart. The data were subdivided into six equally informative distance classes, each one containing from 30 to 34 comparisons: Class limits are shown in Table 4.7a, b and c. Moran's *I* was then calculated for the $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length alleles for each distance class (Table 4.3a, b and c). The formula employed was:

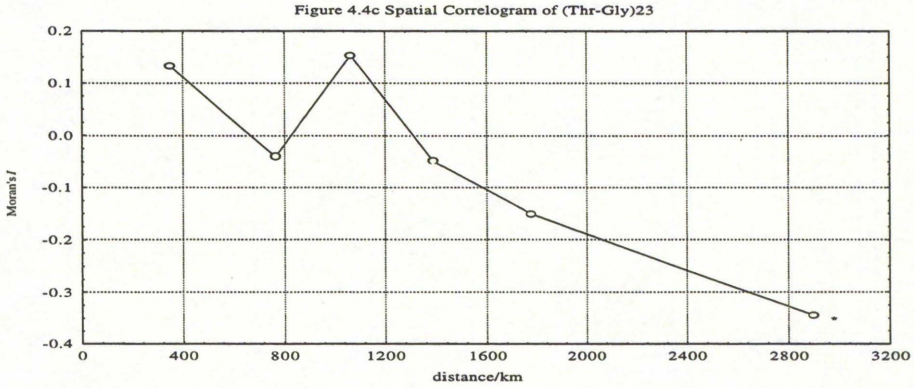
$$I = N\Sigma\Sigma w_{ij}(q_i - q)(q_j - q) / W\Sigma(q_i - q)^2 ,$$

where q_i and q_j are the frequencies of the allele of interest at the *i*th and *j*th locality, q is their mean across the N localities (20 in this case), w_{ij} is equal to 1 for all the pairs of localities falling in the distance class studied and equal to 0 for all other pairs, and W is the sum of all w_{ij} values in that distance class. (See Appendix 4.1 for a working example).

Moran's *I* tends to 1 when the gene frequencies are similar (i.e. depart from the mean in the same direction); it tends to -1 when the gene frequencies are dissimilar, and is expected to be equal to $-1/(N-1)$, under a random hypothesis. A cline is said to be present when Moran's *I* values decrease continuously from significantly positive to significantly negative with an increase in distance between populations (Sokal, 1979). A steady decline of Moran's *I* values was observed for the $(Thr\text{-}Gly)_{20}$ allele from highly positive within ≈350 km to highly negative at large distances (Table 4.7b, Fig. 4.4b). The plots of autocorrelation coefficients versus distance are referred to as correlograms (Figs. 4.4a, b and c). Significance testing of the spatial correlograms was carried out using the Bonferroni procedure as suggested by Oden (1984). This test defines the 'p-value' of the test. If m is the minimum (individual) p-value in a k-class correlogram, then the Bonferroni 'p-value' is km. For example in Table 4.7a, the $(Thr\text{-}Gly)_{17}$ correlogram has 6 distance classes, so $k = 6$. The minimum p for an individual class is 0.086, therefore the Bonferroni approximation is $6 \times 0.086 = 0.516$. The overall trend however was not significant for allele $(Thr\text{-}Gly)_{20}$, as assessed by the Bonferroni test, $p = 0.083$ (Oden, 1984) or for either of the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{23}$ alleles (Table 4.7a and c). Moran's *I* for the $(Thr\text{-}Gly)_{17}$ was non-significant for all distance classes (Table 4.7a), however from inspecting Fig. 4.4a the value of Moran's *I* is increasing up until the distance class 1065-1388







Figures 4.4 a,b and c. Spatial correlograms of the $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length alleles. (* significant Moran's *I*)

Km. The $(Thr\text{-}Gly)_{23}$ variant was observed to have a significantly negative Moran's *I* for the largest distance class although there was no complementary significant positive value. Therefore, there was no significant clinal pattern for the $(Thr\text{-}Gly)_{23}$ frequencies across eastern Australia (Table 4.7c).

The Royaltey-Astrachan-Sokal non-parametric test of departure from random geographic variation (Royaltey *et al.*, 1975) was also applied. The 20 localities were connected by means of a Delaunay graph (Brassel and Reif, 1979) (Fig. 4.1), which is a triangulation network connecting all neighbouring localities together. All the external points are connected if considered related in evolutionary terms and then subsequently subdivided by connecting adjacent localities. The frequencies of the $(Thr\text{-}Gly)_{17}$ allele were ranked among localities. The rank differences between the frequencies of pairs of adjacent localities, called edge lengths, were calculated as was the expected edge length with a correction for continuity. The deviations from the observed mean edge length from the expected mean were then examined with a t-test.

t_s= (observed mean edge length-expected mean edge length $\pm$ correction for continuity)/(standard deviation of mean edge length)

A Student's *t* of 0.93, with infinite degrees of freedom, was observed for the $(Thr\text{-}Gly)_{17}$ allele frequencies, which agreed with the null hypothesis of random variation of this allele. The $(Thr\text{-}Gly)_{20}$ allele frequencies, when calculated in the same way gave a $t = -2.55$, which is significant beyond the 0.02 level. The negative *t* value shows that departure from randomness results from an underlying clinal pattern of allele frequencies (Royaltey *et al.*, 1975). The frequencies of the $(Thr\text{-}Gly)_{23}$ length allele were also treated in the same way and gave a Student's *t* of -2.18 which is significant beyond the 0.05 level, again with infinite degrees of freedom. (A working example of this test is shown in Appendix 4.2).

TABLE 4.7a, b and c. Spatial autocorrelation statistics. Table showing the distance classes used in kilometres, the number of comparisons contained within each class, Moran's *I*, and its relative significance for the $(Thr\text{-}Gly)_{17}$ $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length alleles, respectively.

a $(Thr\text{-}Gly)_{17}$ Overall correlogram significance (Bonferroni approximation) 0.516

DISTANCE CLASSES/KM				
MINIMUM	MAXIMUM	PAIRS	MORANS *I*	p value
0.0	347.9	31	-0.258	0.099
347.9	765.2	34	0.008	0.346
765.2	1065.8	30	0.032	0.297
1065.8	1388.2	31	-0.238	0.124
1388.2	1778.3	32	0.166	0.086
1778.3	2898.3	32	-0.037	0.455

b $(Thr\text{-}Gly)_{20}$ Overall correlogram significance (Bonferroni approximation) 0.093

DISTANCE CLASSES/KM				
MINIMUM	MAXIMUM	PAIRS	MORANS *I*	p value
0.0	347.9	31	0.292	0.015*
347.9	765.2	34	0.194	0.053
765.2	1065.8	30	-0.128	0.319
1065.8	1388.2	31	-0.276	0.082
1388.2	1778.3	32	-0.117	0.345
1778.3	2898.3	32	-0.298	0.039*

c $(Thr\text{-}Gly)_{23}$ Overall correlogram significance (Bonferroni approximation) 0.099

DISTANCE CLASSES/KM				
MINIMUM	MAXIMUM	PAIRS	MORANS *I*	p value
0.0	347.9	31	0.132	0.119
347.9	7 65.2	34	-0.040	0.465
765.2	1065.8	30	0.153	0.093
1065.8	1388.2	31	-0.049	0.491
1388.2	1778.3	32	-0.151	0.266
1778.3	2898.3	32	-0.344	0.017*

The frequencies of the Thr-Gly length alleles were correlated with the yearly mean wet bulb temperature taken at 3pm from the nearest field station (Gentilli, 1971), and a significant correlation was found for the $(Thr\text{-}Gly)_{20}$ variant (Table 4.8). Therefore with an increase in temperature the frequency of $(Thr\text{-}Gly)_{20}$ decreases, implying that flies which carry this length allele are particularly well adapted to cooler, more variable climates. The frequency of the $(Thr\text{-}Gly)_{23}$, when correlated with the same temperature variable, narrowly missed significance.

Table 4.8 Pearson product moment correlations of the frequency of the Thr-Gly variants with respect to mean yearly wet bulb temperature ^{o}C (*, $p<0.05$)

Allele	*r*	p value
$(Thr\text{-}Gly)_{14}$	0.10	0.673
$(Thr\text{-}Gly)_{17}$	- 0.12	0.626
$(Thr\text{-}Gly)_{18}$	0.14	0.548
$(Thr\text{-}Gly)_{20}$	- 0.54	0.014*
$(Thr\text{-}Gly)_{21}$	0.41	0.073
$(Thr\text{-}Gly)_{22}$	0.17	0.473
$(Thr\text{-}Gly)_{23}$	- 0.44	0.053
$(Thr\text{-}Gly)_{24}$	- 0.05	0.844

The observation of a significant correlation of mean yearly temperature with the frequency of the $(Thr\text{-}Gly)_{20}$ carrier across Australia is different to that found for the European $(Thr\text{-}Gly)_{20}$ allele which was not significantly correlated. In addition, the European $(Thr\text{-}Gly)_{17}$ allele frequency was correlated with temperature, whereas the Australian $(Thr\text{-}Gly)_{17}$ variant was not. This is consistent with the stronger cline seen in Europe for the $(Thr\text{-}Gly)_{17}$ allele compared to the $(Thr\text{-}Gly)_{20}$ carrier (Costa *et al.*, 1992), and also the absence of a $(Thr\text{-}Gly)_{17}$ and presence of a $(Thr\text{-}Gly)_{20}$ variant cline in Australia (this chapter).

Comparing the European and Australian clines for the $(Thr\text{-}Gly)_{20}$ allele

When comparing the two clines (Fig. 4.5) it is interesting to note the difference in gradient between them, the European cline being the steeper of the two. This could be due to the much greater periods of time since *D. melanogaster* colonised Europe compared to Australia.

4.4 DISCUSSION

The finding of a significant correlation of European 'mean Thr-Gly length' with mean yearly temperature could suggest an effect of this environmental parameter in maintaining the cline. Perhaps the most relevant observation is that the frequency of the $(Thr\text{-}Gly)_{17}$ allele increases with an increase in temperature in the summer months. This implies that the $(Thr\text{-}Gly)_{17}$ variant may be particularly adapted to warmer climates. These new findings together with the predominance of the $(Thr\text{-}Gly)_{17}$ allele in southern Europe (Costa *et al.*, 1992) suggests to us that the shorter of the two most common length alleles may be more suited to these increased temperatures.

The $(Thr\text{-}Gly)_{17}$ allele is probably a recent derivative of the $(Thr\text{-}Gly)_{20}$ (Rosato *et al.*, 1996). This view is supported by the finding that the frequency of the $(Thr\text{-}Gly)_{17}$ length allele is very low in Africa (C. Pasetto pers. comm.), the ancestral home of *D. melanogaster*. The presence of the European $(Thr\text{-}Gly)_{17}$ allele cline could then be seen as a result of an initial population admixture, where once *D. melanogaster* had begun to colonise Europe, 10, 000 years ago, the more recently derived $(Thr\text{-}Gly)_{17}$ expanded its range in Europe and started to compete with the $(Thr\text{-}Gly)_{20}$ variant. If it was favoured in the warmer climates of southern Europe and north Africa this admixture would be reinforced by selection. Given the effective population size of *D. melanogaster* of approximately 10^6 (Aquadro, 1992), only weak selection would be required on the $(Thr\text{-}Gly)_{17}$ carrier to generate these geographical patterns (see also final chapter). It is also of interest that no correlation with temperature was observed with the

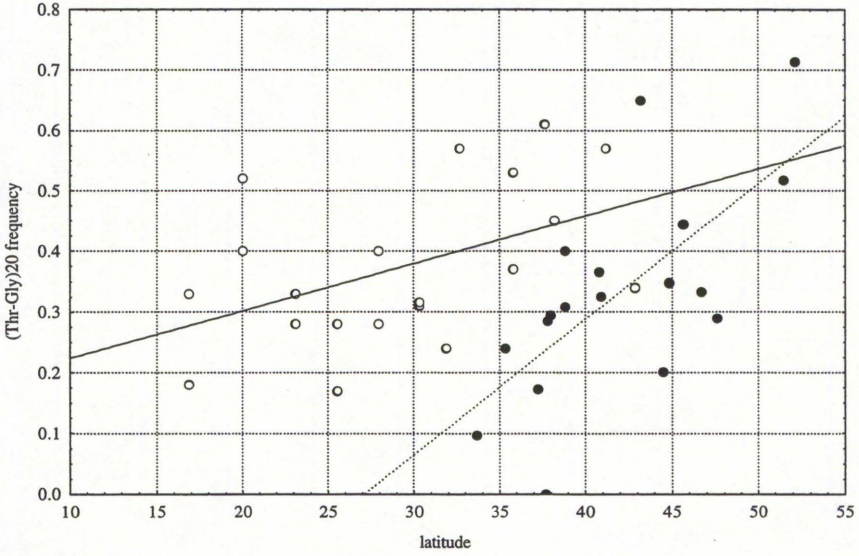


Figure 4.5 Comparing the plots of the European (•) and Australian (o) clines for the $(Thr\text{-}Gly)_{20}$ length allele

$(Thr\text{-}Gly)_{20}$. If we assume that temperature may be a relevant factor in European population dynamics for these Thr-Gly alleles, then perhaps the $(Thr\text{-}Gly)_{20}$ allele frequencies are altogether less sensitive to temperature changes. This will be explored in chapter 5.

In the Australian samples, there was also good evidence for clinal variation in Thr-Gly length, providing additional support for the hypothesis that these clines in repeat length may be maintained by selective gradients. The decline of spatial autocorrelation with distance for the $(Thr\text{-}Gly)_{20}$ allele, and the result of the Royaltey-Astrachan-Sokal test agree in suggesting this variant is geographically structured along an approximately north-south axis of Australia. Also the weak correlations and spatial structure for the $(Thr\text{-}Gly)_{23}$ allele suggest a higher prevalence of the $(Thr\text{-}Gly)_{23}$ in the north compared to the south of Australia. However, there is no compelling evidence for a robust cline with this variant. The observation of high $(Thr\text{-}Gly)_{17}$ allele frequencies overall, and particularly in the north, when compared to the corresponding $(Thr\text{-}Gly)_{20}$ allele frequencies is interesting and could be explained by founder effects followed by selective migration. Moreover, the absence of a complementary cline in the $(Thr\text{-}Gly)_{17}$ allele might possibly be explained if selection favours only the $(Thr\text{-}Gly)_{20}$ allele in cooler climates, thereby leaving warmer environments open to exploitation by any other allele. However a more plausible reason is based on history as Australia's recent colonisation compared to that of Europe, means that *D. melanogaster* migration has only been widespread for two hundred years at most. If weak selection is acting on the Thr-Gly repeat, as suggested by the linkage disequilibrium study of Rosato *et al* (ms. submitted), then longer periods of time may be required for the stabilisation of any spatial pattern. This Australian colonisation is thought to have occurred from Africa, Asia and Europe (David and Capy, 1988) and one can imagine that along the eastern coast the colonisation and subsequent migration by humans and their fly commensals will have contributed to a constant admixture of the founder Thr-Gly variants. It might therefore appear surprising that any cline in Thr-Gly length was discovered at all on this continent, particularly given Rosato *et al's* (ms. submitted) result which reveals that weak selection is probably acting on the array.

What would provide unequivocal evidence for a selective explanation of the clines? The discovery of parallel patterns of geographical variation for shared polymorphisms in different sibling species, as shown by Anderson and Oakeshott (1984) for the *Est6* and phosphoglucomutase (*Pgm*) polymorphisms of *D. melanogaster* and *D. simulans*, would go some way to meet this requirement. They observed the two species sharing the same latitudinal correlated patterns of *Est6* and *Pgm* allele frequencies in Australasia, North America and Europe. However, the molecular polymorphism of the Thr-Gly region in natural European populations of *D. simulans* showed different and fewer rarer length alleles than those found in *D. melanogaster*, and they were homogenously distributed across Europe and North Africa (Rosato *et al.*, 1994). Why does the same gene show such a radically different pattern of geographical variation in *D. melanogaster*, where a robust cline is observed in Thr-Gly length (Costa *et al*,. 1992 and this chapter) compared to *D. simulans'* apparent spatial homogeneity? Different selective regimes might be maintaining the cline in *D. melanogaster* and the geographically homogenous distribution of the haplotype frequencies in *D. simulans*. In these regimes the effective population size (N_e) of the two species becomes important. In a larger population such as *D. simulans* (Aquadro, 1992), genetic drift is less important as an evolutionary factor, but selection against slightly deleterious mutants is more efficient due to the larger N_e and therefore any 'nearly neutral' mutations (Ohta, 1996) are more visible to selection and are either eliminated or fixed. Thus *D. melanogaster* will maintain more 'nearly neutral' mutations than *D. simulans*. Although both species were genetically versatile to colonise temperate regions, the greater 'potentially adaptiv' genetic variation in *D. melanogaster* may have permitted a spatial differentiation in Europe in contrast with *D. simulans*, where any variability was lost or fixed.

In conclusion, finding that the frequency of a major Thr-Gly variant shows a strong latitudinal cline in Australia provides further evidence that natural selection might be operating to maintain this polymorphism. Although allele frequency gradients on such large geographical scales can be generated by other process, such as range expansion, a classical explanation for a latitudinal cline is adaptive response to climatic variation. Climate related selection is thought

to be responsible for latitudinal clines at other loci in *D.melanogaster*, as in phenotypic traits such as thorax length and wing area (James and Partridge, 1995: R. Stanley pers. comm.). Although no correlation with temperature was found for the population frequencies of European Thr-Gly length variants (Costa *et al.*, 1992), a more sophisticated analysis using a more refined data set has found a significant correlation with temperature and average Thr-Gly length for each population. The significant positive correlations observed between the frequency of the $(Thr\text{-}Gly)_{17}$ variant and temperature in the hotter months of the year in Europe further adds to the overall impression that selection could be an important factor in the European cline. James and Partridge (1995), found that all temperature variables were negatively correlated with latitude in Australia, regardless of the time of day or seasons used to estimate them. The correlation observed with the frequency of $(Thr\text{-}Gly)_{20}$ alleles in Australia and temperature suggests that this cline is also temperature related.

It is interesting to observe that the great majority of the Thr-Gly variants analysed so far from European, North African and Australian populations of *D. melanogaster* differ in length by multiples of three Thr-Gly pairs. The exceptions account for approximately 10% of the total sample from both continents (114 in 1126). If the Thr-Gly mutation rate via slippage or unequal crossovers is quite high, then one would perhaps expect a less 'periodic' distribution of length variation. However structural studies of the Thr-Gly dipeptide have revealed that a hexapeptide ($(Thr\text{-}Gly)_3$) forms a stable beta turn which shows significant flexibility at high temperatures (Castiglione-Morelli *et al.*, 1995). If the $(Thr\text{-}Gly)_3$ represents a conformational monomer for the Thr-Gly array, then perhaps variants whose lengths fall outside the 14, 17, 20, 23 periodic series have less stable Thr-Gly tracts, and this might provide a structural basis by which to understand how thermal challenges could effect Thr-Gly frequencies. In any case, it appears that perhaps selection, in addition to historical factors in Thr-Gly allele evolution (Rosato *et al.*, 1996), has produced these continental patterns of polymorphism.

CHAPTER 5

5.1 INTRODUCTION

In chapter 3 a sequence analysis of the length polymorphism in the Thr-Gly repeat region of the *per* gene in *D. melanogaster* populations from both Europe and Eastern Australia was reported. The presence of such a polymorphism, the latitudinal cline (Costa *et al.*, 1992) in the two common variants and the similar inverse cline for one of the frequent length variants in Eastern Australia (see chapter 4) prompts the question as to whether natural selection is acting on this polymorphism and whether any adaptive behavioural differences could be associated with the different Thr-Gly length variants, particularly the more common, $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles. A preliminary experiment (Peixoto, PhD thesis 1993a) of this kind was carried out on one $(Thr\text{-}Gly)_{14}$ length variant and two $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles. A significant difference in freerunning period was found for the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{17}$ length alleles at different temperatures but not for the $(Thr\text{-}Gly)_{20}$ length allele. This preliminary result is consistent with studies by Ewer *et al.* (1990), who showed that the period of the clock in per^{01} transformants carrying a Thr-Gly deleted *per* gene became temperature sensitive. Consequently I have performed an extensive behavioural analysis of many Thr-Gly length variants at two different temperatures to determine whether there exists any phenotypic differences between them involving temperature compensation.

5.2 METHODS

Single fly DNA extraction, PCR, purification, direct DNA sequencing and automated DNA sequencing were carried out as described in chapter 2.

The phenotypic analysis presented in this chapter is taken from flies collected in different localities in Europe and North Africa. The sample from Pietrastornina (PI) was obtained in October 1989, while the samples from Casablanca (CAS) in North Africa, Leiden (LE) in the Netherlands and Rethimnon (RET) in Crete were collected in September-October 1989. Flies from Cognac (CO) in France were collected in October-November 1991. The samples from Conselve (CON) in Italy were collected in October

1992. A population from North Wooton, Somerset, UK was also collected in October 1994. For more details see chapters 3, 4 and Costa *et al.* (1991; 1992).

For the behavioural analysis, males carrying either the $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17\text{-}2}$, $(Thr\text{-}Gly)_{20}$, $(Thr\text{-}Gly)_{21}$, $(Thr\text{-}Gly)_{23}$ and $(Thr\text{-}Gly)_{24}$ *per* length variants were crossed to attached-*X* *ywf* females (stock obtained from Dr R. Greenspan). This enabled the *X* chromosome from the parental male to be represented in all of the male progeny. Thirty-two of these attached-*X* lines were made from single males carrying various Thr-Gly length variants from laboratory stocks isolated originally from the natural European populations (see above). The additional five attached-*X* lines from Cognac used in the behavioural analysis each carry an *X* chromosome originally from homozygous Thr-Gly length variant stocks that were isolated from isofemale lines collected in Cognac. Males from a laboratory stock of Canton-S, (Ohio, U.S.A) were also crossed to attached-*X* females. The behavioural experiments were carried out in constant darkness at 18°C and 29°C as described in chapter 2.

Statistical analysis of the data was carried out using the Minitab version 8 software and Statistica (StatSoft) for windows. In the ANOVAs to be presented the number of planned comparisons usually involved all the individual genotypes at the two temperatures. If for example, four genotypes are examined at two temperatures, the Genotype and Temperature main effects have four degrees of freedom between them, three for Genotype, one for Temperature, and also three for the Temperature Genotype interaction, allowing seven individual planned comparisons to be made (Sokal and Rohlf, 1981). Any further comparisons require adjustments to the acceptable probability level (α) used to reject the null hypothesis of the effect, in this case $\alpha = 0.05$. However, when performing many of these tests there is a probability that some are significant by chance. By setting the acceptable α to 0.05 the probability that a or more significant tests out of n performed were significant by chance was calculated using the binomial expansion, and when appropriate, this probability is also given.

5.3. RESULTS

Table 5.1 and Table 5.2 show the mean freerunning periods obtained at 18°C and 29°C for flies carrying $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17\text{-}2}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$, $(Thr\text{-}Gly)_{21}$, $(Thr\text{-}Gly)_{23}$ or $(Thr\text{-}Gly)_{24}$, variants analysed by autocorrelation and spectral procedures respectively. Thirtyeight lines from different origins were tested, four $(Thr\text{-}Gly)_{14}$, one $(Thr\text{-}Gly)_{17\text{-}2}$, twelve $(Thr\text{-}Gly)_{17}$, eleven $(Thr\text{-}Gly)_{20}$, three $(Thr\text{-}Gly)_{21}$, six $(Thr\text{-}Gly)_{23}$ alleles (including Canton-S which is American) and one $(Thr\text{-}Gly)_{24}$ carrier

The *a priori* hypothesis was that flies carrying different Thr-Gly length variants would show differences in the temperature sensitivity of their clock, and further, this difference might contribute to an explanation of the geographical distribution of the Thr-Gly length alleles. The descriptive statistics based on autocorrelation and spectral analyses are given for each line, in Table 5.1 and 5.2 respectively

5.3.1 Global analysis of variance (ANOVA) of Thr-Gly length variants pooled within genotype

As an initial analysis a global two way ANOVA was performed with Genotype (7 levels) and Temperature (2 levels) as the main effects, followed by individual planned comparisons of each genotype across temperature, for autocorrelation and spectral data respectively.

Autocorrelation

The autocorrelation data gave significant Genotype and Temperature effects but no significant interaction effect, see Table 5.3.

Table 5.3. Two way fixed ANOVA, Autocorrelation (*, $p<0.05$)

Source	MS	d.f	F	p
Genotype(G)	2.87	6	3.25	0.004*
Temperature(T)	11.34	1	12.83	0.000*
G*T	1.62	6	1.84	0.088
Error	0.884	1481		

AUTOCORRELATION (*, p<0.05; MS error, 0.85, df, 1419)

	line	N/S	genotype	mean period @ 18°C	n @ 18°C	SEM+/-	mean period @ 29°C	n @ 29°C	SEM+/-	absolute difference (ABS)	F	p -level	latitude degrees
1	CON	N	14A	23.68	33	0.20	24.48	29	0.10	0.80	11.71	0.001*	45.23
2	CON	N	14B	23.61	28	0.16	24.39	28	0.10	0.78	10.22	0.001*	45.23
3	CO	N	14	23.76	29	0.14	23.20	28	0.14	-0.56	3.23	0.073	45.683
4	LEC	S	14	23.77	13	0.33	23.84	16	0.15	0.07	0.47	0.828	40.35
5	CO	N	15	23.88	12	0.33	24.08	19	0.25	0.20	0.36	0.548	45.683
6	CON	N	17A	23.97	31	0.15	24.45	32	0.12	0.48	4.39	0.036*	45.23
7	CON	N	17B	23.87	15	0.34	23.78	29	0.10	-0.09	0.10	0.756	45.23
8	CON	N	17C	23.32	19	0.24	24.25	24	0.13	0.93	10.95	0.001*	45.23
9	LE	N	17	23.26	21	0.27	23.83	9	0.25	0.57	2.43	0.119	52.16
10	PI.1	S	17	23.42	12	0.27	23.61	14	0.18	0.19	0.28	0.599	40.916
11	PI.9	S	17	23.60	15	0.27	24.25	14	0.16	0.65	3.62	0.057	40.916
12	CAS	S	17	23.31	59	0.12	23.73	40	0.07	0.42	4.92	0.027*	33.66
13	RET.9	S	17	23.81	27	0.22	23.96	12	0.22	0.15	0.23	0.632	35.33
14	RET.2	S	17	24.07	7	0.76	24.60	10	0.22	0.53	1.36	0.244	35.33
15	LEC60	S	17	23.32	19	0.27	24.30	20	0.14	0.98	11.16	0.001*	40.35
16	LEC12	S	17	24.14	28	0.18	24.25	22	0.17	0.11	0.17	0.683	40.35
17	NW6	N	17	23.89	13	0.35	23.95	20	0.12	0.06	0.04	0.842	51.45
18	CO	N	20	23.48	25	0.22	23.67	29	0.15	0.19	0.59	0.443	45.683
19	LE	N	20	23.83	20	0.27	23.92	12	0.19	0.09	0.07	0.785	52.16
20	PI.5	S	20	23.75	4	0.83	23.63	8	0.16	-0.12	0.05	0.824	40.916
21	PI.8	S	20	23.85	10	0.48	24.13	8	0.13	0.28	0.39	0.528	40.916
22	CAS	S	20	23.79	38	0.17	23.40	13	0.19	-0.39	1.70	0.192	33.66
23	RET.4	S	20	24.34	16	0.28	24.68	11	0.28	0.34	0.88	0.348	35.33
24	CON	N	20A	23.42	25	0.22	23.72	30	0.13	0.30	1.42	0.234	45.23
25	CON	N	20B	23.36	22	0.24	24.18	26	0.19	0.82	9.68	0.002*	45.23
26	LEC6	S	20	23.93	15	0.21	24.00	14	0.22	0.07	0.04	0.845	40.35
27	LEC3	S	20	23.46	12	0.32	24.25	14	0.14	0.79	4.79	0.029*	40.35
28	NW1	N	20	23.35	10	0.26	23.56	18	0.16	0.21	0.32	0.571	51.45
29	CO	N	21	23.59	16	0.27	23.41	17	0.30	-0.18	0.32	0.570	45.683
30	CON	N	21A	23.95	26	0.35	24.08	25	0.13	0.13	0.25	0.619	45.23
31	CON	N	21B	23.65	27	0.19	24.05	22	0.16	0.40	2.26	0.133	45.23
32	CO	N	23	23.97	19	0.23	24.73	11	0.14	0.76	4.68	0.031*	45.683
33	CON	N	23A	23.93	30	0.22	24.00	24	0.10	0.07	0.07	0.791	45.23
34	CON	N	23B	24.05	21	0.14	24.00	23	0.12	-0.05	0.03	0.864	45.23
35	Canton.S	--	23	23.81	16	0.23	23.94	19	0.16	0.13	0.17	0.676	U.S.A
36	LEC79	S	23	23.33	12	0.39	23.77	13	0.15	0.44	1.40	0.236	40.35
37	LEC30	S	23	24.19	8	0.28	23.75	10	0.34	-0.44	1.00	0.316	40.35
38	NW5	N	24	23.17	15	0.30	23.47	16	0.14	0.03	0.84	0.361	51.45

TABLE 5.1 Mean free running locomotor activity periods at 18°C and 29°C derived from autocorrelation analysis of European and North African flies carrying different Thr-Gly lengths. Population abbreviations are as follows, Pietrastornina (PI), Casablanca (CAS), Leiden (LE), Lecce (LEC), Rethimnon (RET), Cognac (CO), Conselve (CON), North Wooton (NW). Standard error of the means (SEM), planned comparison results, absolute difference between means, latitudes and an arbitrary grouping (see text) of northern(N) and southern(S) populations for each line are also shown. The Genotype column refers to the Thr-Gly length of the lines A. B and C codes for genotype do not refer to different isolength alleles but different lines within genotype.

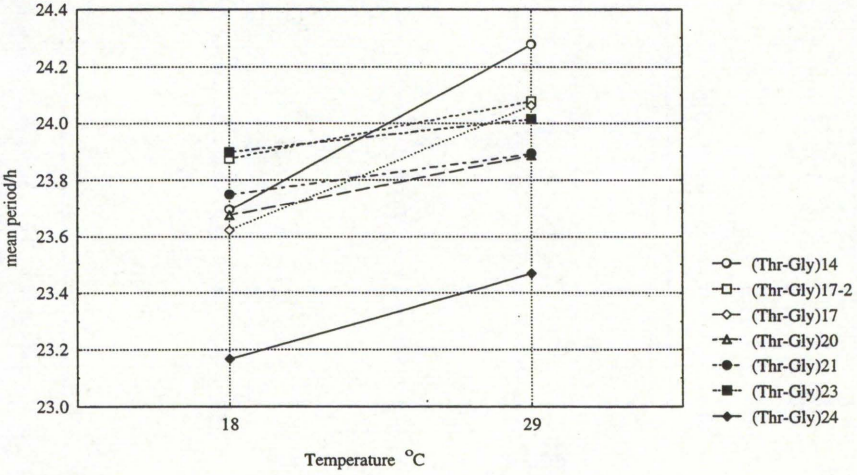
SPECTRAL(*, p<0.05; MS error 0.55, df 1419)

	line	N/S	genotype	mean period @ 18°C	n @ 18°C	SEM+/-	mean period @ 29°C	n @ 29°C	SEM+/-	absolute difference (ABS)	F	p-level	latitude
1	CON	N	14A	23.97	33	0.15	24.12	29	0.12	0.15	0.62	0.430	45.23
2	CON	N	14B	23.43	28	0.16	24.00	28	0.14	0.57	8.18	0.004*	45.23
3	CO	N	14	23.92	29	0.13	24.16	28	0.13	0.24	1.54	0.215	45.683
4	LEC	S	14	23.85	13	0.17	23.69	16	0.17	-0.16	0.31	0.582	40.35
5	CO	N	15	23.92	12	0.20	23.65	19	0.16	-0.27	0.98	0.322	45.68
6	CON	N	17A	23.94	31	0.13	24.55	32	0.15	0.61	10.55	0.001*	45.23
7	CON	N	17B	23.31	15	0.19	23.71	29	0.15	0.40	2.91	0.088	45.23
8	CON	N	17C	23.66	19	0.29	23.91	24	0.20	0.25	1.24	0.266	45.23
9	LE	N	17	23.49	21	0.15	23.78	9	0.36	0.29	0.96	0.326	52.16
10	PI.1	S	17	23.75	12	0.10	23.91	14	0.25	0.16	0.29	0.585	40.92
11	PI.9	S	17	23.99	15	0.16	23.81	14	0.22	-0.18	0.43	0.511	40.92
12	CAS	S	17	23.21	59	0.08	23.09	40	0.17	-0.12	0.57	0.452	33.66
13	RET.9	S	17	23.84	27	0.19	23.94	12	0.15	0.10	0.14	0.705	35.33
14	RET.2	S	17	24.20	7	0.32	24.37	10	0.20	0.17	0.22	0.642	35.33
15	LEC60	S	17	24.18	19	0.17	24.31	20	0.17	0.13	0.33	0.565	40.35
16	LEC12	S	17	24.13	28	0.12	24.14	22	0.17	0.01	0.00	0.981	40.35
17	NW6	N	17	24.48	13	0.14	23.84	20	0.14	-0.64	5.69	0.017*	51.45
18	CO	N	20	23.70	25	0.11	23.39	29	0.18	-0.31	2.38	0.123	45.68
19	LE	N	20	23.81	20	0.11	23.91	12	0.30	0.10	0.13	0.717	52.16
20	PI.5	S	20	23.37	4	0.42	23.61	8	0.22	0.24	0.29	0.586	40.92
21	PI.8	S	20	24.01	10	0.18	24.28	8	0.16	0.27	0.59	0.441	40.92
22	CAS	S	20	23.69	38	0.08	23.12	13	0.15	-0.57	5.66	0.017*	33.66
23	RET.4	S	20	23.81	16	0.18	24.20	11	0.29	0.39	1.81	0.178	35.33
24	CON	N	20A	23.47	25	0.13	23.28	30	0.14	-0.19	0.90	0.344	45.23
25	CON	N	20B	23.72	22	0.17	23.79	26	0.14	0.07	0.11	0.742	45.23
26	LEC6	S	20	23.90	15	0.13	24.40	14	0.19	0.50	3.32	0.068	40.35
27	LEC3	S	20	23.88	12	0.14	24.27	14	0.12	0.39	1.78	0.182	40.35
28	NW1	N	20	23.73	10	0.23	23.20	18	0.17	-0.53	3.17	0.075	51.45
29	CO	N	21	23.47	16	0.21	23.85	17	0.20	0.38	2.18	0.140	45.68
30	CON	N	21A	23.81	26	0.16	23.93	25	0.13	0.12	0.34	0.562	45.23
31	CON	N	21B	23.71	27	0.10	23.96	22	0.16	0.25	1.31	0.253	45.23
32	CO	N	23	24.38	19	0.19	24.78	11	0.15	0.40	1.98	0.159	45.68
33	CON	N	23A	23.99	30	0.15	23.75	24	0.14	-0.24	1.36	0.243	45.23
34	CON	N	23B	24.03	21	0.14	24.03	23	0.14	-0.00	0.00	0.996	45.23
35	canton.s		23	24.00	16	0.19	23.19	19	0.21	-0.81	9.70	0.002*	U.S.A
36	LEC79	S	23	23.95	12	0.17	23.49	13	0.19	-0.46	2.42	0.120	40.35
37	LEC30	S	23	24.18	8	0.17	23.70	10	0.34	-0.48	1.85	0.173	40.35
38	NW5	N	24	23.77	15	0.15	23.11	16	0.14	-0.66	6.20	0.013*	51.45

TABLE 5.2 Mean free running locomotor activity periods at 18°C and 29°C derived from spectral analysis of European and North African flies carrying different Thr-Gly lengths. Population abbreviations are as follows, Pietrastornina (PI), Casablanca (CAS), Lecce (LEC), Leiden (LE), Rethimnon (RET), Cognac (CO), Conselve (CON), North Wooton (NW). Standard error of the means (SEM), planned comparison results, absolute difference between means, latitudes and an arbitrary grouping (see text) of northern(N) and southern(S) populations for each line are also shown. The Genotype column refers to the Thr-Gly length of the lines A. B and C codes for genotype do not refer to different isolength alleles but different lines within genotype.

The planned comparisons of pooled Thr-Gly length variants (Appendix 5.1) showed the $(Thr\text{-}Gly)_{14}$ ($F = 19.6$, $p << 0.01$), $(Thr\text{-}Gly)_{17}$ ($F = 28.1$, $p << 0.01$) and $(Thr\text{-}Gly)_{20}$ ($F = 4.8$, $p < 0.05$) length alleles to have significantly different length periods between temperatures. The mean freerunning locomotor activity periods for each genotype at the two temperatures are shown in Fig. 5.1. All the length variants are increasing their period with an increase in temperature and it can be seen that the largest differences between temperature are found in the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{17}$ alleles. It is also quite noticeable that the new and rare European $(Thr\text{-}Gly)_{24}$ variant has a circadian rhythm that is considerably shorter than that of the other length variants.

Fig. 5.1 Mean periods of the European Thr-Gly length variants
2-way interaction (autocorrelation)
F(6,1481)=1.84; p<0.09

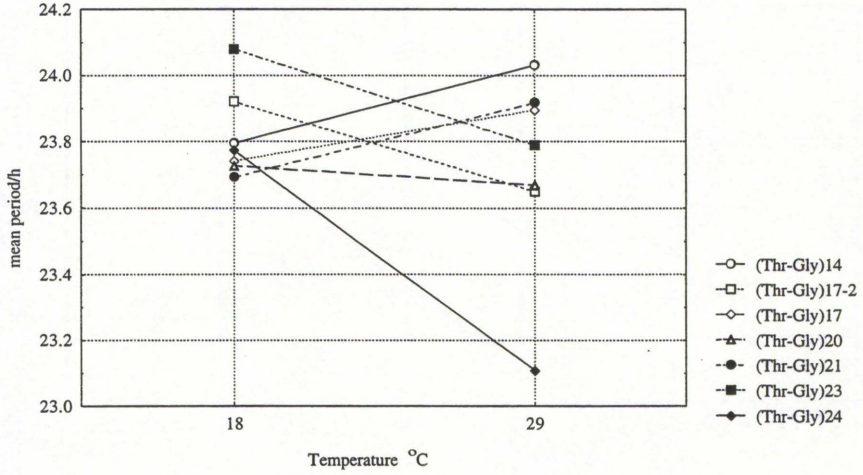


Spectral

For spectral data a significant Genotype effect ($F = 3.79$, $p << 0.01$) and Genotype Temperature (G*T) interaction ($F = 4.16$, $P << 0.01$) was observed (Appendix 5.2). Planned comparisons of genotype period length at different temperatures pooled within genotype (Appendix 5.3) showed significant differences for the $(Thr\text{-}Gly)_{14}$ ($F = 4.51$, $p < 0.05$), $(Thr\text{-}Gly)_{17}$ ($F = 4.75$, $p < 0.05$), $(Thr\text{-}Gly)_{23}$ ($F = 6.82$, $p < 0.01$) and $(Thr\text{-}Gly)_{24}$ ($F = 5.4$, $p < 0.05$) length variants. The plot of the mean locomotor activity periods for each genotype is shown in Fig. 5.2. In contrast with the autocorrelation data, the Thr-Gly length variants

either increase or decrease their period with respect to temperature. However, all show marked deviations in period length with the higher temperature apart from the $(Thr\text{-}Gly)_{20}$.

Fig. 5.2 Mean periods of the European Thr-Gly length variants
2-way interaction (spectral)
F(6,1481)=4.16; p<0.0004



Overall this global analysis shows that the Thr-Gly length variants are behaving differently at 18°C and 29°C when individuals are pooled across lines and into their appropriate genotypes. There is compelling evidence from inspection of the means (Appendix 5.3), that of the major variants, the $(Thr\text{-}Gly)_{20}$ shows the most efficient temperature compensation, and this is largely supported by the various ANOVAs.

In order to dissect this further at the global level, independent ANOVAs were performed on the common $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles and on two extreme length alleles the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$, which might provide the best chance of a genotype related difference.

5.3.2 Unnested ANOVAs on $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles.

The $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ allele lines were pooled within genotype to see whether a Temperature Genotype (T*G) interaction occurred globally.

Autocorrelation

A two way ANOVA was performed and a significant Temperature effect was observed. The Temperature and Genotype (T*G) interaction effect narrowly missed significance (Table 5.4).

Table 5.4. Unnested ANOVA on $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles. (*, $p<0.05$)

AUTOCORRELATION				
EFFECT	MS	df	F	p
GENOTYPE(G)	0.86	1	0.97	0.323
TEMP(T)	23.18	1	26.37	0.000*
T*G	2.87	1	3.26	0.071
ERROR	0.879	888		

When followed by planned significant tests both the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ showed significantly different periods at the two temperatures ($F = 28.27$, $p < 0.001$; $F = 4.83$, $p < 0.05$, respectively) suggesting the two were behaving in a similar way. However on inspecting the plot of the means this difference can be seen quantitatively (Fig. 5.3a), the $(Thr\text{-}Gly)_{20}$ having a smaller difference between temperatures for period length than the $(Thr\text{-}Gly)_{17}$.

Spectral

A significant Genotype effect ($F = 4.67$, $P<0.05$) was observed when a two-way ANOVA was performed on spectral data. The Temperature Genotype (T*G) interaction narrowly missed significance ($p = 0.057$, Appendix 5.4). This suggested that the two alleles were behaving in dissimilar ways. This is illustrated more clearly in Fig. 5.3b, where not only do the two alleles differ in the direction of compensation at 29°C, but it is also evident that the $(Thr\text{-}Gly)_{20}$ maintains its locomotor activity period rather better than the $(Thr\text{-}Gly)_{17}$. Planned comparisons revealed a significant difference between period length for the $(Thr\text{-}Gly)_{17}$ ($F = 4.42$, $p<0.05$) and not the $(Thr\text{-}Gly)_{20}$ ($F = 0.49$, $p>0.1$).

Overall this pooled analysis which focuses more closely on the two most common Thr-Gly length variants shows an interesting genotypic effect in temperature compensation.

Fig.5.3a Mean periods for the (Thr-Gly)17 and (Thr-Gly)20 alleles
2-way interaction (autocorrelation)
F(1,888)=3.26; p<0.0713

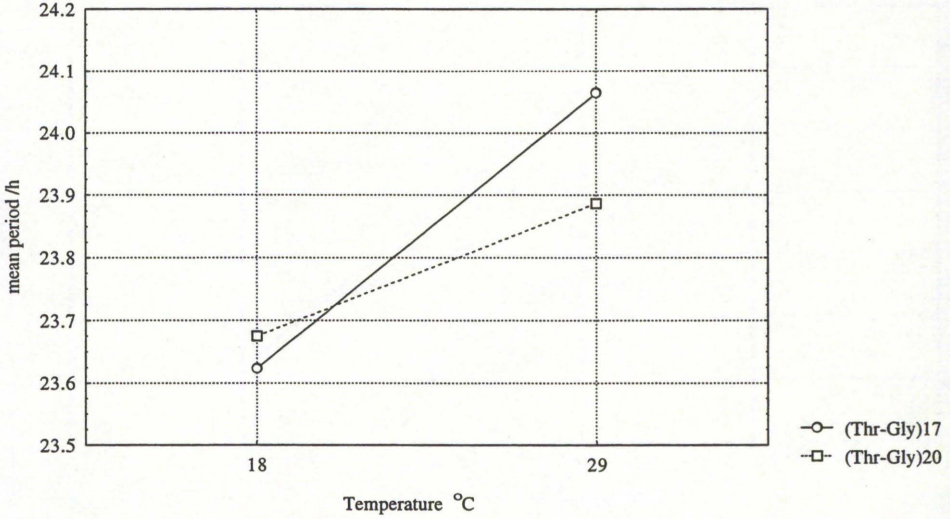


Fig. 5.3b Mean periods of the (Thr-Gly)17and (Thr-Gly)20 alleles
2-way interaction (spectral)
F(1,888)=3.62; p<0.0574

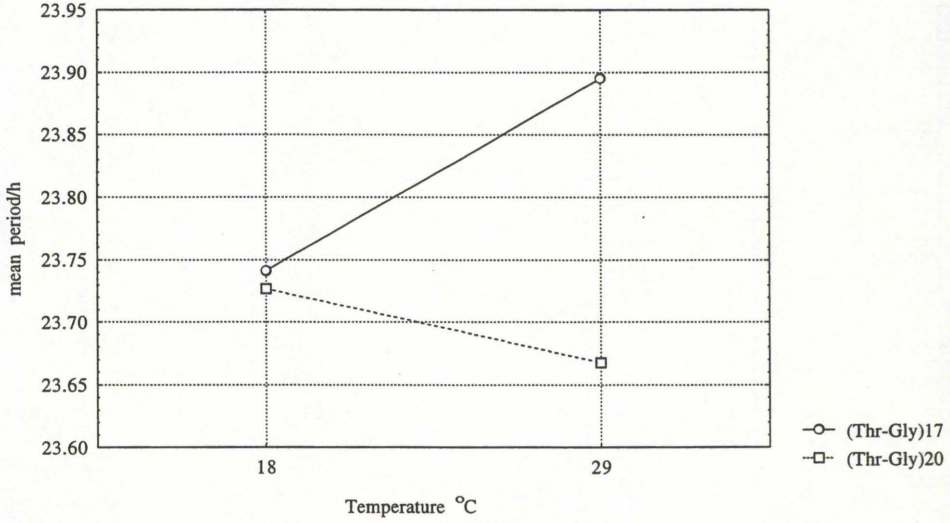


Figure 5.3a and b. Mean periods of the two common length variants the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ at different temperatures, for autocorrelation and spectral respectively.

5.3.3 Unnested ANOVA on $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ length alleles.

The $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ lines were similarly pooled within genotype and the American Canton-S line was excluded in this operation.

Autocorrelation

A two way analysis of variance was performed and a significant Temperature effect and Temperature and Genotype (T*G) was revealed indicating that these two genotypes when pooled across lines differed significantly with respect to temperature, see Table 5.5.

Table 5.5. Unnested ANOVA on $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ length alleles. (*, $p<0.05$)

AUTOCORRELATION				
EFFECT	MS	df	F	p
GENOTYPE(G)	0.02	1	0.0265	0.871
TEMP(T)	11.47	1	15.099	0.0001*
T*G	4.984	1	6.561	0.0108*
ERROR	0.759	371		

When followed by planned comparisons the $(Thr\text{-}Gly)_{14}$ variants showed a significant difference between temperature ($F = 23.8$, $p<<0.001$), whereas the $(Thr\text{-}Gly)_{23}$ length variants did not ($F = 0.98$, $p> 0.1$). This marked difference in temperature compensation can be seen in Fig. 5.4a, where it is apparent that the $(Thr\text{-}Gly)_{23}$ carrier compensates its period to a greater extent than the $(Thr\text{-}Gly)_{14}$ allele.

Spectral

On repeating this analysis with the spectral periods, again there was a significant Temperature Genotype interaction (T*G) ($F = 6.81$, $p < 0.01$, Appendix 5.5). As before, the planned comparison tests were significant for the $(Thr\text{-}Gly)_{14}$ variants at the two temperatures ($F= 5.02$, $p<0.05$), but not for the $(Thr\text{-}Gly)_{23}$ variants ($F = 2.21$, $p>0.1$).

Therefore a Genotype Temperature interaction (T*G) was observed in the pooled data, suggesting that the shorter Thr-Gly repeat lengths confer a greater sensitivity to temperature than the longer Thr-Gly lengths, even though Fig.5.4b gives an impression that both variants are temperature sensitive and in opposite directions.

Fig 5.4a Mean periods for the (Thr-Gly)14 and (Thr-Gly)23 alleles
2-way interaction (autocorrelation)
F(1,371)=6.56; p<0.0108

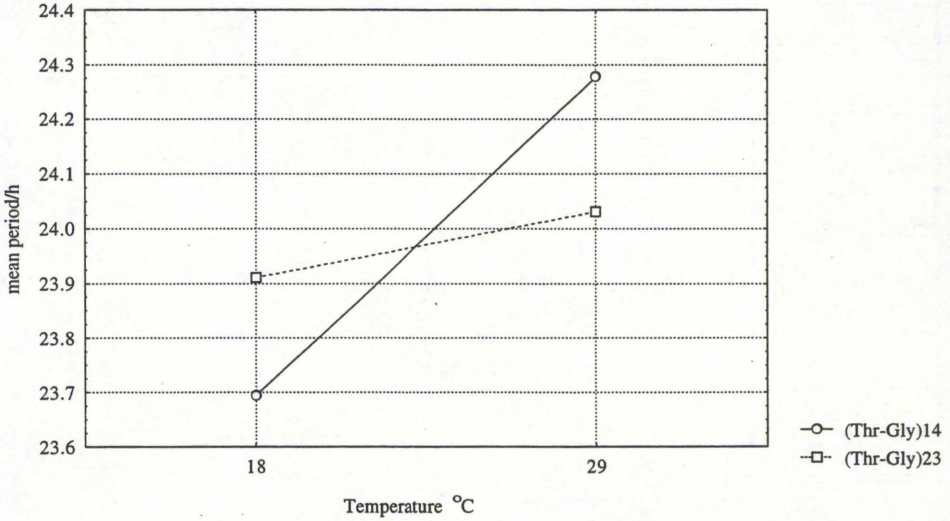
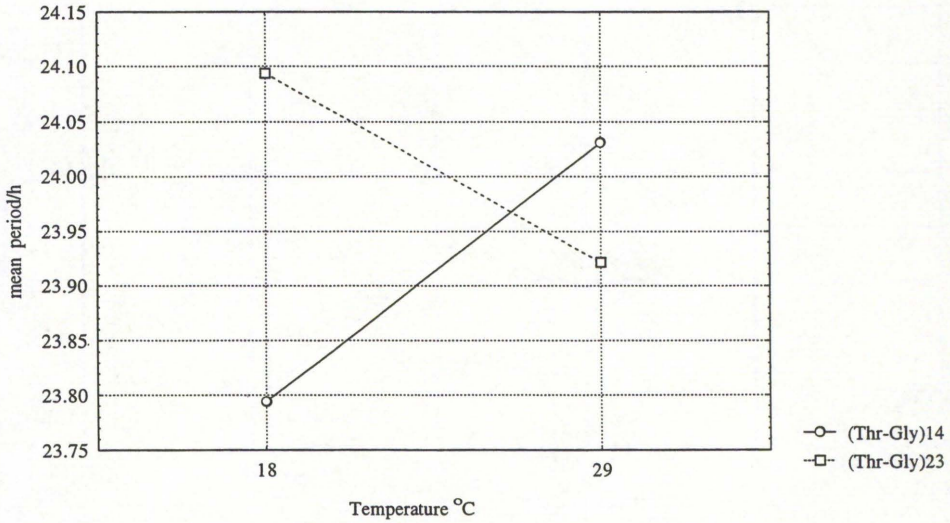


Fig. 5.4b Mean periods for the (Thr-Gly)14 and (Thr-Gly)23 alleles
2-way interaction (spectral)
F(1,371)=6.81; p<0.0094



Figures 5.4a and b. Mean periods of the two extreme length variants the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$, at different temperatures, for autocorrelation and spectral respectively.

5.3.4 ANOVA for all 38 lines.

A two way ANOVA, fixed model, with Lines (38 levels) and Temperature (2 levels) as the main effects was performed, followed by individual planned comparisons (see F statistic in Tables 5.1 and 5.2).

Autocorrelation

The autocorrelation results gave significant Line and Temperature effects (p= 0.000 for both) but no significant interaction (Table 5.6).

Table 5.6 Two way fixed ANOVA, Autocorrelation (*, p<0.05)

Source	MS	d.f	F	p
TEMPERATURE(T)	27.16	1	32.13	0.000*
LINE(L)	2.53	37	2.99	0.000*
T*L	1.11	37	1.31	0.104
ERROR	0.85	1419		

The planned comparisons (Table 5.1) showed 2 out of 4 $(Thr\text{-}Gly)_{14}$ allele lines to be significantly different in period between 18°C and 29°C (CON14A and CON14B), as were 4 out of the 12 lines tested which carried a $(Thr\text{-}Gly)_{17}$ allele (CON17A, CON17C, CAS4 and LEC60). Only 2 out of the 11 lines analysed carrying a $(Thr\text{-}Gly)_{20}$ length allele showed a significant difference (CON20B and LEC3) and a single $(Thr\text{-}Gly)_{23}$ repeat line (CO23), out of the possible 6, including Canton-S. The probability that nine or more tests are significant out of 38, assuming a probability of 0.05 that a test might be significant by chance, is 0.00008. Clearly it appears that these effects are real.

Spectral

The spectral results showed a significant Line effect ($F = 7.05$, $p<<0.01$), and a significant Line and Temperature interaction (T*L, $F = 2.21$, $p<<0.01$, Appendix 5.6 and Table 5.2). Overall, spectral planned comparisons gave a more conservative result with a smaller number of lines significantly different between the two temperatures, see Table 5.2. Only one $(Thr\text{-}Gly)_{14}$ line was significantly different (CON14B) between temperature, and 2 $(Thr\text{-}Gly)_{17}$ lines (CON17A and NW6). In addition 3 $(Thr\text{-}Gly)_{20}$ alleles were significantly

Fig. 5.5a Mean locomotor activity periods of all 38 lines
2-way interaction (autocorrelation)
F(37,1419)=1.31; p<0.1035

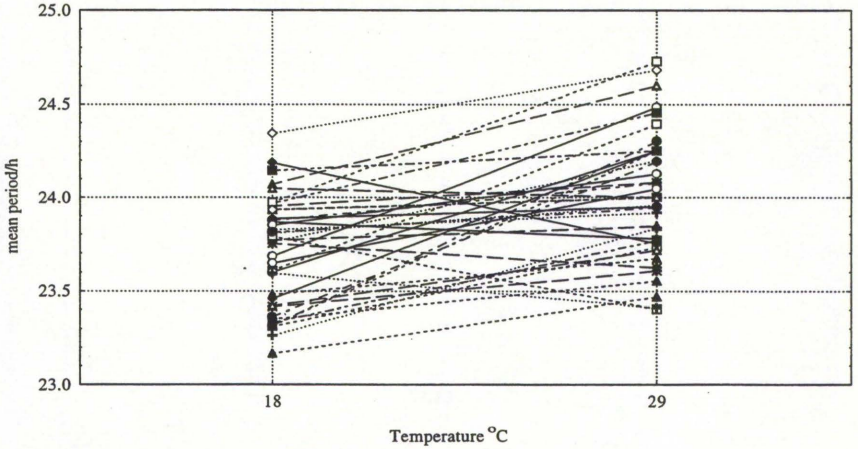


Fig. 5.5b Mean locomtor activity periods of all 38 lines
2-way interaction
F(37,1419)=2.21; p<0.0000

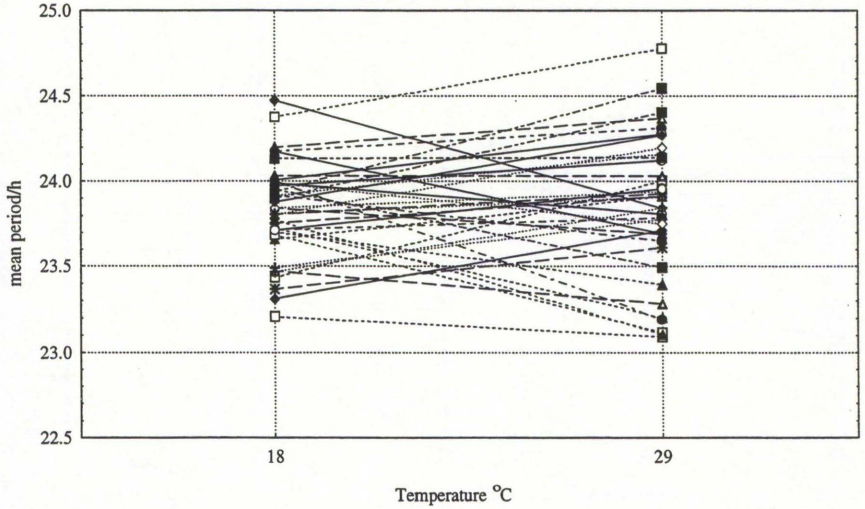


Figure 5.5a and b. Mean locomotor activity periods of all 38 lines plotted across temperature for autocorrelation and spectral analyses respectively.

different (CAS10, LEC6 and LEC3), only one of which had shown a difference when analysed with autocorrelation. Again a single $(Thr\text{-}Gly)_{23}$ line Canton-S (CS) showed a statistically significant difference in period length at the two temperatures. It is interesting to note that this line shows a downstream coding polymorphism not seen in any other of the European lines (Rosato *et al.*, 1996). Lastly, the only other line showing a significant difference of period length between 18°C and 29°C was the single $(Thr\text{-}Gly)_{24}$ length allele (NW5). The probability of eight or more planned comparisons being significant by chance is 0.00049, supporting again the validity of these results.

When plotted as means for both spectral and autocorrelation data it is apparent that the lines behave differently across the temperatures (Fig. 5.5a and b). However the majority of lines are increasing their period with respect to temperature based on autocorrelation, compared to those analysed spectrally whose periods both increase and decrease, giving rise to the significant interaction term in the ANOVA.

The differences in results from the two analyses, autocorrelation and spectral, could possibly be due to the coarseness of the methods employed. Where autocorrelation was used, a period value can only be obtained to the nearest 30 minutes, whereas spectral analysis was able to resolve periods down to about 6-10 minutes (see chapter 2) .

In conclusion the overall ANOVA's showed that the different lines behave differently across temperatures. Autocorrelation shows that more lines of the shorter variants ($(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{17}$) tend to show differences between the two temperatures whereas the results from the spectral analysis are more conservative.

5.3.5 ANOVA on individual genotypes

Isolength Thr-Gly repeat alleles can also be treated as a single group, so the periods between lines within each genotype were analysed. The Canton-S line was excluded from this analysis since it did not originate from Europe.

Autocorrelation

The most relevant source of variation is the Temperature effect which is highly significant for the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{17}$ and marginally significant for $(Thr\text{-}Gly)_{20}$ length alleles (Table 5.7). No significant Temperature Line interaction (T*L) was observed. The Line effect is significant for the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length

TABLE 5.7 AUTOCORRELATION

REPEAT LENGTH ALLELES

	$(Thr\text{-}Gly)_{14}$				$(Thr\text{-}Gly)_{17}$				$(Thr\text{-}Gly)_{20}$				$(Thr\text{-}Gly)_{21}$				$(Thr\text{-}Gly)_{23}$			
EFFECT	MS	df	F	p	MS	df	F	p	MS	df	F	p	MS	df	F	p	MS	df	F	p
TEMP(T)	12.7	1	19.2	0.000*	17.5	1	21.9	0.000*	3.9	1	4.66	0.032*	0.4	1	0.31	0.58	0.86	1	1.03	0.313
LINE(L)	0.5	3	0.76	0.52	3.1	11	3.88	0.000*	2.5	10	2.92	0.002*	2.6	2	1.95	0.146	2.13	4	2.55	0.041*
T*L	1.2	3	1.74	0.161	1.1	11	1.39	0.172	1.1	10	1.26	0.252	0.8	2	0.61	0.545	1.53	4	1.53	0.196
ERROR	0.66	196			0.8	488			0.85	358			1.4	127			0.83	161		
TOTAL		203				511				379				132				170		

TABLE 5.8 SPECTRAL

REPEAT LENGTH ALLELES

	$(Thr\text{-}Gly)_{14}$				$(Thr\text{-}Gly)_{17}$				$(Thr\text{-}Gly)_{20}$				$(Thr\text{-}Gly)_{21}$				$(Thr\text{-}Gly)_{23}$			
EFFECT	MS	df	F	p	MS	df	F	p	MS	df	F	p	MS	df	F	p	MS	df	F	p
TEMP(T)	1.89	1	3.51	0.063	1.0	1	1.55	0214	0.09	1	0.19	0.662	1.97	1	3.68	0.057	0.88	1	1.7	0.195
LINE(L)	1.58	3	2.93	0.035*	7.09	11	10.9	0.000*	2.85	10	6.29	0.000*	0.49	2	0.92	0.402	3.03	4	5.81	0.000*
T*L	0.91	3	1.69	0.170	1.04	11	1.61	0.0.93	1.1	10	2.42	0.009*	0.17	2	0.32	0.727	0.87	4	1.67	0.158
ERROR	0.54	196			0.65	488			0.45	358			0.54	127			0.52	161		
TOTAL		203				511				379				132				170		

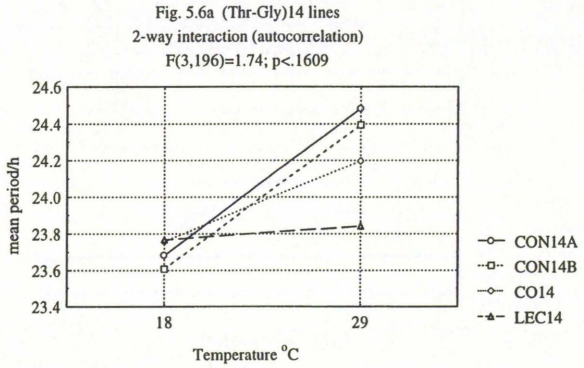
Table 5.7 and 5.8 ANOVA on individual genotypes for autocorrelation and spectral periods respectively.
*, $p < 0.05$.

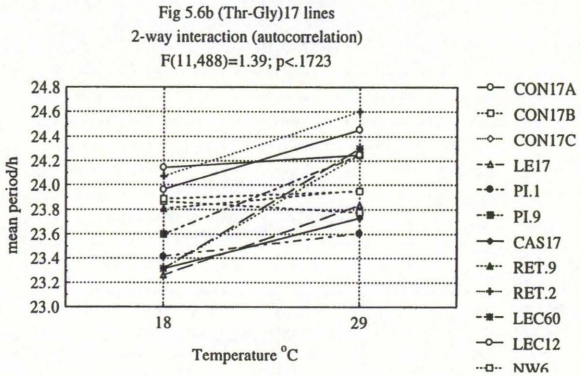
repeat. Figures 5.6a - e show the mean locomotor activity periods for all the lines within each genotype. In Fig 5.6a, all the $(Thr\text{-}Gly)_{14}$ lines appear to differ significantly between temperatures apart from LEC14 which appears to maintain its period length. The majority of the $(Thr\text{-}Gly)_{17}$ lines (Fig. 5.6b) appear to increase their period length with temperature, however there are a few that are able to maintain their period between temperatures. In general the $(Thr\text{-}Gly)_{20}$ lines (Fig. 5.6c) cover a smaller range than the $(Thr\text{-}Gly)_{17}$ lines, apart from RET4 which showed elevated period lengths for both temperatures compared to the other lines. Also the majority of the lines appear to have homogenous period lengths for both temperatures, only a few differing to any great extent. Two out of three of the $(Thr\text{-}Gly)_{21}$ lines (Fig. 5.6d) show poor temperature compensation, whereas two out of five $(Thr\text{-}Gly)_{23}$ lines (Fig. 5.6e) show good temperature compensation. These figures and this analysis shows that there are line differences for temperature compensation within genotypes, but it is still generally apparent that the $(Thr\text{-}Gly)_{20}$ allele shows better overall temperature compensation, followed by the $(Thr\text{-}Gly)_{23}$ variant, while the shorter length alleles the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{17}$, and the infrequent $(Thr\text{-}Gly)_{21}$ variant show greater temperature variation in period length.

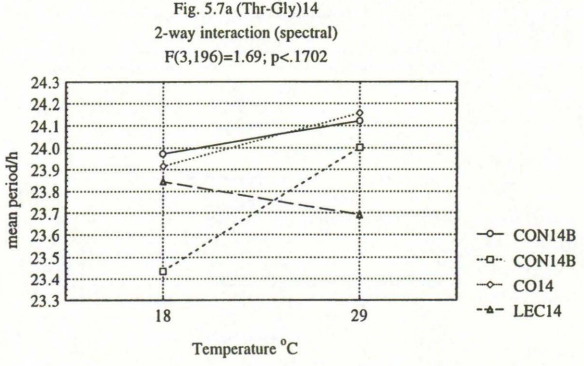
Spectral

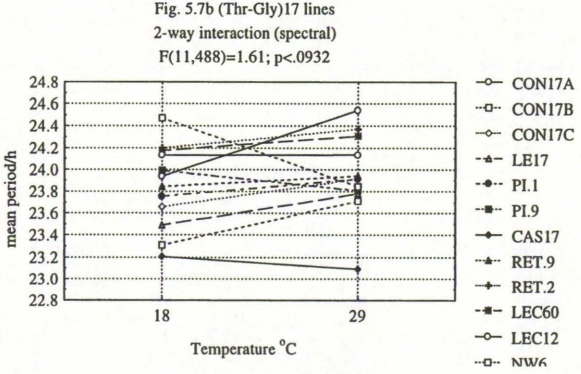
A Temperature effect was not seen for any genotype, however Line effects were observed for $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$, and $(Thr\text{-}Gly)_{23}$ length repeats. A significant Temperature and Line (T*L) interaction was observed with the $(Thr\text{-}Gly)_{20}$ length alleles (Table 5.8).

Figs. 5.7a to e show the mean locomotor activity periods for lines within each genotype tested. Generally the $(Thr\text{-}Gly)_{14}$ lines are poor in temperature compensation, however, two lines CON14B and CO14 appear to maintain their period length reasonably well across temperature compared to the other two lines (Fig. 5.7a). Again, as with autocorrelation data the period length range covered by all the $(Thr\text{-}Gly)_{17}$ lines is greater than that covered by the $(Thr\text{-}Gly)_{20}$, particularly at 18°C (Fig. 5.7b). The majority of $(Thr\text{-}Gly)_{17}$ lines are increasing their period length with respect to temperature. The $(Thr\text{-}Gly)_{20}$ lines appear to both lengthen and shorten when tested at the higher temperature (Fig. 5.7c). The rare $(Thr\text{-}Gly)_{21}$ allele also shows poor temperature compensation, Fig. 5.7d. Some of the $(Thr\text{-}Gly)_{23}$ lines appear to stabilise their period quite well, however three lines do not CO23, LEC79 and LEC30 (Fig. 5.7d).









Figs. 5.6a,b, 5.7a and b Mean locomotor activity periods for lines within each genotype

Fig. 5.6c (Thr-Gly)20 lines
2-way interaction (autocorrelation)
F(10,358)=1.26; p<.2523

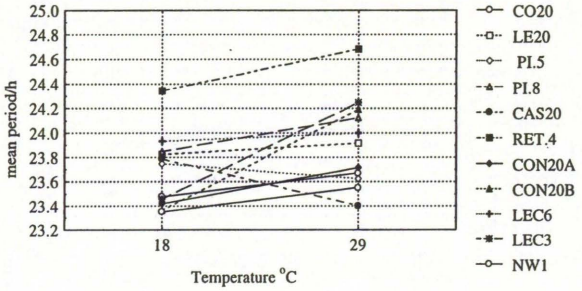


Fig. 5.6d (Thr-Gly)21 lines
2-way interaction (autocorrelation)
F(2,127)=.61; p<.5450

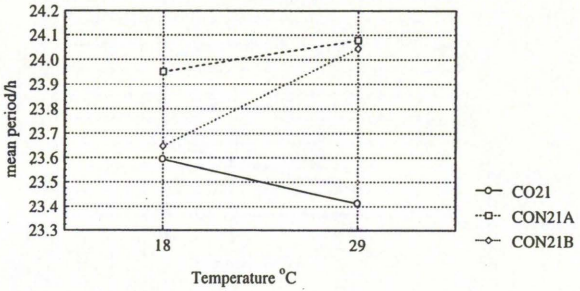


Fig. 5.7c (Thr-Gly)20 lines
2-way interaction (spectral)
F(10,358)=2.42; p<.0086

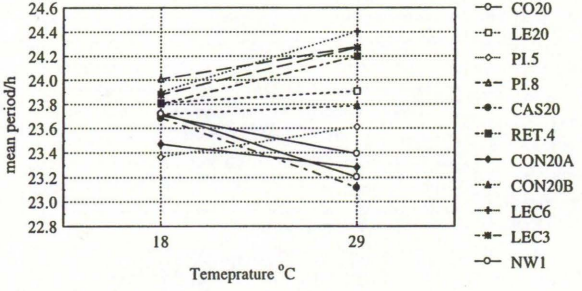
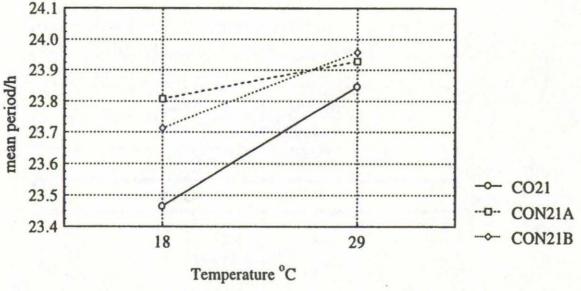
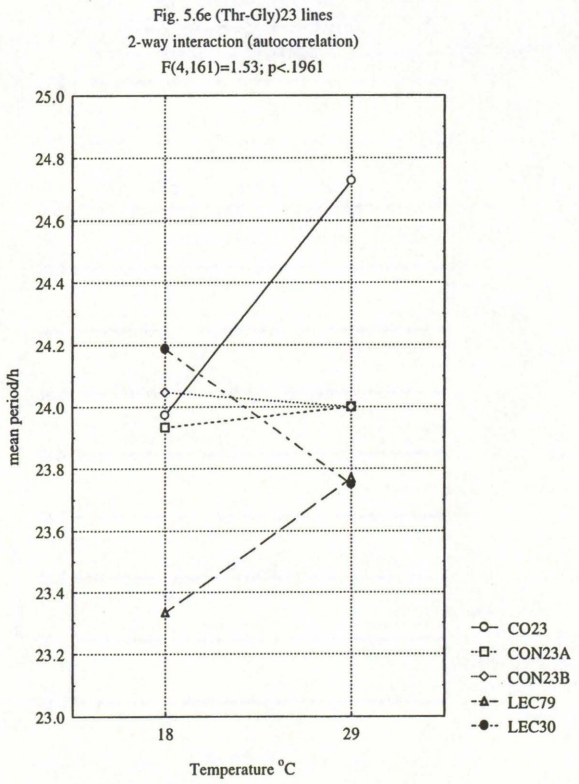
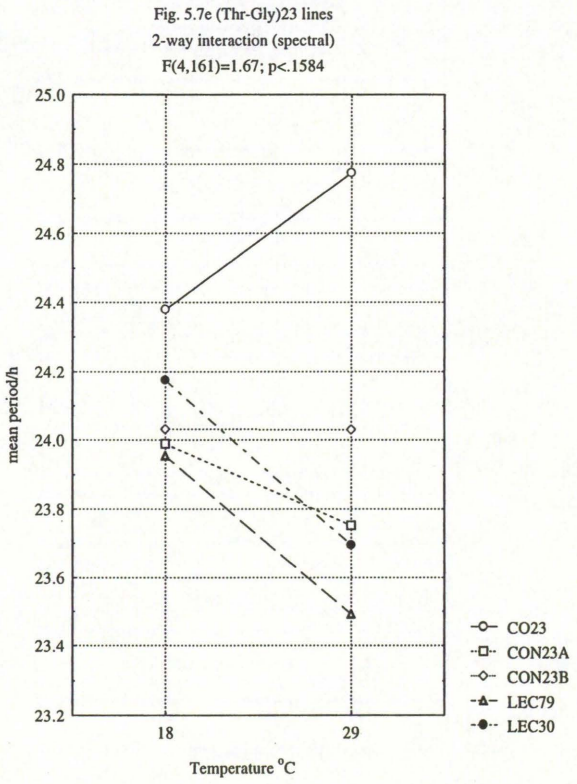


Fig. 5.7d (Thr-Gly)21 lines
2-way interaction (spectral)
F(2,127)=.32; p<.7265



Figs. 5.6c,d, 5.7c and d Mean plots continued.





Figs. 5.6e and 5.7e Mean plots continued

In summary the absence of a significant Temperature Line interaction for each genotype, apart from the $(Thr\text{-}Gly)_{20}$ variants, suggests the lines are generally behaving similarly within a genotype. However the significant Line effects seen for the majority of genotypes does suggest that the lines have different periods irrespective of temperature. The strong temperature effect with autocorrelation also indicates that the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{17}$ alleles are less well compensated than the longer alleles.

5.3.6 Pooling the Thr-Gly lines

The pooled mean periods across lines were plotted as a function of temperature, as illustrated in Figs. 5.8a and 5.8b for autocorrelation and spectral data respectively. For example the $(Thr\text{-}Gly)_{17}$ allele gives a mean based on the 12 values for each line at each temperature. These unweighted means are therefore different from the means based on individual data (Fig. 5.1 and 5.2).

Autocorrelation

The autocorrelation periods show the general trend (Fig. 5.8a.) that the shorter length variants have a larger difference between the periods at 18°C and 29°C. However, the periods of the longer length variants, $(Thr\text{-}Gly)_{20}$ or more repeats, are not as sensitive to temperature changes, with the exception of the $(Thr\text{-}Gly)_{24}$ variant, where only one line was examined. All the mean periods derived from autocorrelation procedures appear to lengthen as the temperatures rises.

Spectral

For spectral data (Fig. 5.8b) the periods are observed to lengthen or shorten in an allele specific manner. Generally the $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ variants, whose length differs by $(Thr\text{-}Gly)_3$, are less sensitive to temperature than the rarer alleles, $(Thr\text{-}Gly)_{17\text{-}2}$, $(Thr\text{-}Gly)_{21}$ and $(Thr\text{-}Gly)_{24}$ alleles. Fig. 5.8b shows that the period of the $(Thr\text{-}Gly)_{20}$ is the best compensated of the longer length variants. When considering the $(Thr\text{-}Gly)_{17}$ length allele we observe a period nearer 24 h for both autocorrelation and spectral analyses at 29°C whereas the $(Thr\text{-}Gly)_{20}$ allele shows an appreciable deviation (approx. 0.3 h) from 24 hours.

Both these figures show the same phenomenon as Fig. 5.1 and Fig. 5.2, the shorter Thr-Gly length variants being more sensitive to temperature than the longer ones.

Figure 5.8a Mean periods (autocorrelation)

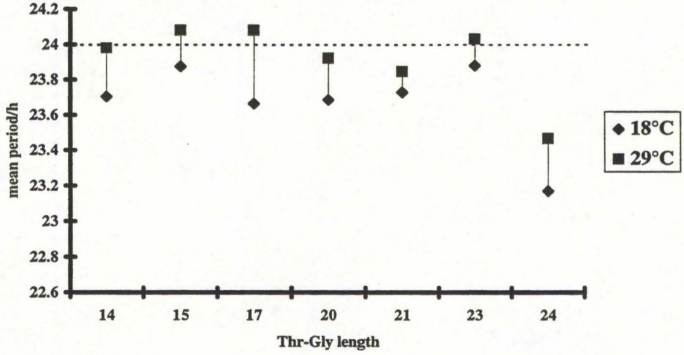


Figure 5.8b Mean periods (spectral)

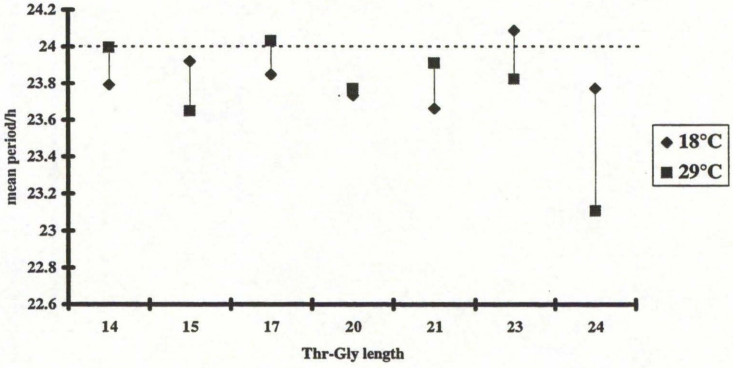


Figure 5.8a and 5.8b. Mean locomotor activity periods for Thr-Gly length variants found within Europe (Note the $(Thr\text{-}Gly)_{15}$ variant is a $(Thr\text{-}Gly)_{17}$ lacking the $(Thr\text{-}Gly)_2$ repeat downstream of the main repeat, $(Thr\text{-}Gly)_{17\text{-}2}$).

Comparison of the two analyses shows that the sample size differences within each line do not disproportionately affect the conclusions.

5.3.7 Pooling mean temperature differences in period across lines

Plotting the difference in mean periods at the two temperatures for the different Thr-Gly variants reveals some interesting observations (Fig. 5.9a and 5.9b). These differences were examined by calculating the mean absolute difference between the two periods at different temperatures (mean period @ 29°C-mean period @ 18°C) within each genotype by using the line means, then adding +1 to change any negative values to positive ones. Thus the mean absolute difference for the $(Thr\text{-}Gly)_{17}$ allele is based on 12 lines and the $(Thr\text{-}Gly)_{14}$ variant on four lines etc...

Autocorrelation

Differences in mean autocorrelation periods (Fig. 5.9a) showed no significant correlation (Pearson product moment, $r = -0.61$, $p > 0.1$, $df = 5$), indicating no simple linear relationship between temperature compensation and Thr-Gly length. When the rare alleles were excluded, the $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ variants revealed a highly significant correlation between temperature compensation and Thr-Gly length ($r= 0.99$, $p<0.02$, $df = 2$).

Spectral

Similarly, with spectral analysis (Fig. 5.9b) no significant correlation coefficient was recorded ($r= -0.57$, $p >0.1$, $df = 5$), but again when excluding all the rarer alleles, the $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length variants generate a significant correlation ($r= 0.98$, $p<0.02$, $df = 2$). It is evident from Fig 5.9a and particularly 5.9b, that the common variants of 14, 17, 20 and 23 Thr-Gly pairs lie appreciably closer to the linear regression line than the rarer alleles. This also suggests that the $(Thr\text{-}Gly)_3$ repeat unit is integral to the relationship between temperature compensation and Thr-Gly length.

Figure 5.9a Mean period difference (autocorrelation)

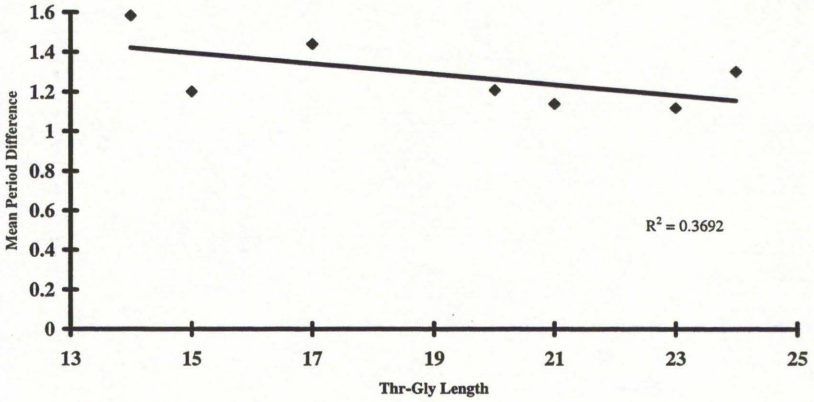


Figure 5.9b Mean period difference (spectral)

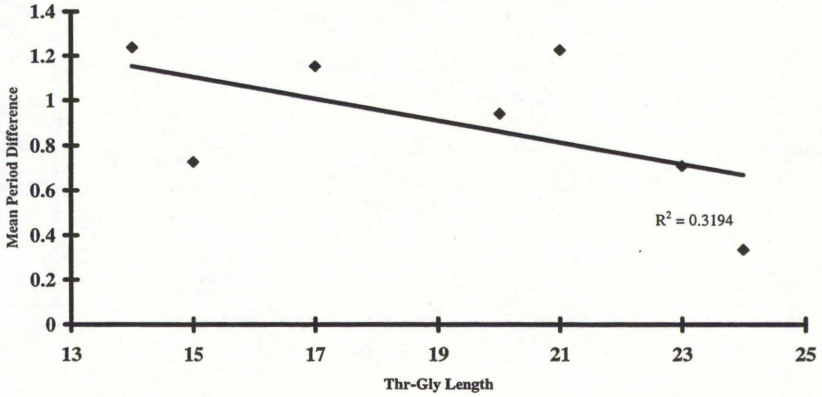


Figure 5.9a and 5.9b. Difference in mean locomotor activity period between temperatures for each European Thr-Gly length variant.

5.3.8 Period differences between temperatures for individual lines

When using the temperature difference in period between individual lines for each genotype, this correlation coefficient appears less convincing (r= -0.19, p>0.1, df, 36; r =-0.31, p>0.05, df, 36), as illustrated by Fig. 5.10a and 5.10b, both for autocorrelation and spectral data. However, the general trend is still noticeable, whereby an increase in Thr-Gly length results in a decrease in mean difference of periods across temperatures. Also for both autocorrelation and spectral data it is noticeable that the majority of absolute differences for the $(Thr\text{-}Gly)_{17}$ lines are all in a positive direction, while these differences for the $(Thr\text{-}Gly)_{20}$ alleles are distributed evenly both negatively and positively. This is also true for the $(Thr\text{-}Gly)_{23}$ length variant. The observation that the $(Thr\text{-}Gly)_{17}$ variants all increase their locomotor activity period length with temperature, whereas the $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length alleles increase or decrease their period with an increase in temperature is intriguing and will be discussed later.

5.3.9 Nested ANOVA on $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles

It was clear from the previous ANOVA's that Lines had to be taken into consideration. To do this a nested design had to be implemented since each line did not recur with each genotype. Unfortunately nesting the lines was extremely difficult as the number of lines for each genotype was unbalanced and therefore the model on which such an ANOVA is based is incomplete. It was decided therefore to compare the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles within a balanced nested ANOVA. This meant reducing the number of lines of $(Thr\text{-}Gly)_{17}$ length alleles from 12 to 11 lines. RET2, the line having the smallest sample size was excluded (see Table 5.1, 5.2, RET2, line 14).

Autocorrelation

The Mean Squares, degrees of freedom, F and p values are shown in Table 5.9 for all effects. With the autocorrelation data, Temperature and Line effects were significant. However, there was no significant interaction between Temperature and Genotype (T*G) which would have been expected if the two genotypes had shown a differences between temperatures. The *a priori* comparisons performed by line (Appendix 5.7) gave essentially identical results to those presented in Table 5.1, even though the effects are tested against the error variance of only the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles. Four $(Thr\text{-}Gly)_{17}$ lines had

Figure 5.10a Period difference for individual lines (autocorrelation)

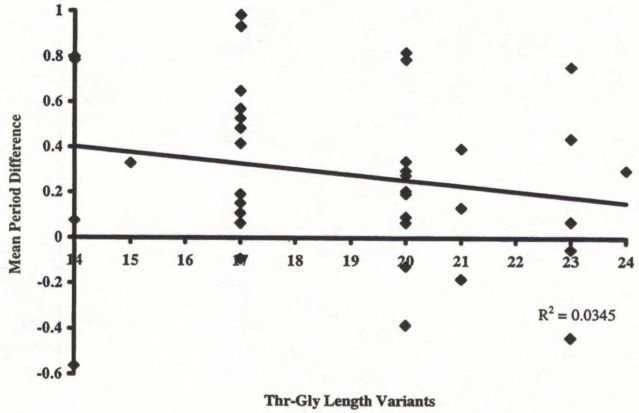


Figure 5.10b Period difference for individual lines(spectral)

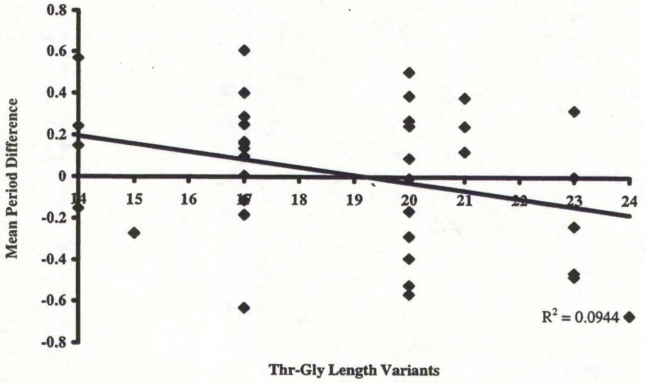


Figure 5.10a and 5.10b Period difference means for each line of Thr-Gly length variants.

significantly different circadian periods between temperatures, CON17A, CON17C, CAS17 and LEC60, also PI9 narrowly missed significance. Two $(Thr\text{-}Gly)_{20}$ lines were also significantly different (CON20B and LEC3) between temperatures

TABLE 5.9 Nested ANOVA on $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles (*, $p<0.05$)

AUTOCORRELATION				
EFFECT	MS	df	F	p
GENOTYPE(G)	0.11	1	0.13	0.718
TEMP(T)	17.39	1	21.66	0.000*
LINE(L)	2.75	20	3.42	0.000*
T*G	1.25	1	1.56	0.212
L*T	1.15	20	1.43	0.100
ERROR	0.80	831		

Spectral

With spectral data (Appendix 5.8) the Line effect was again significant as was the Line Temperature interaction (L*T), that is to say that lines were acting differently irrespective of being nested within Genotype. Again no significant interaction was seen between Temperature and Genotype (T*G), but the Genotype effect narrowly failed to reach significance. The planned comparisons (Appendix 5.9) gave almost identical results to those shown in Table 5.2. A single $(Thr\text{-}Gly)_{20}$ line (CAS20) was significantly different between 18°C and 29°C, two $(Thr\text{-}Gly)_{17}$ lines (CON17A and NW6) were also significantly different between temperatures.

Two planned comparisons were performed for the line that was excluded for the nested analysis (RET2) in a background of all the other eleven $(Thr\text{-}Gly)_{17}$ lines. For both autocorrelation and spectral periods there was no significant difference between the periods at 18°C and 29°C.

These nested ANOVA's revealed a marginal Genotype effect (spectral), and both analyses suggested a slightly less efficient temperature compensation in $(Thr\text{-}Gly)_{17}$ variants versus $(Thr\text{-}Gly)_{20}$ variants.

5.3.10 Nested ANOVA on $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ length alleles.

It was also of interest to directly compare the other 2 most common alleles, $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$, since they were the shortest and longest alleles to be found, and perhaps created the best opportunity for observing any phenotypic difference. Again, to do this type of analysis, lines had to be nested within genotype and because of unequal number of lines between the two genotypes the $(Thr\text{-}Gly)_{23}$ line that had the smallest sample size, LEC30 was excluded (see Table 5.1, 5.2, LEC30, line 37). Canton-S was again excluded because only European lines were being compared

Autocorrelation

When using periods from the autocorrelation procedure there was a significant Temperature and Line effect, see Table 5.10. However no interactions were significant and this was borne out by inspecting the mean plot of the temperature genotype interaction (Fig. 5.11a) where both genotypes increased period length with an increase in temperature.

Table 5.10 Nested ANOVA on $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ length alleles. (*, $p<0.05$)

AUTOCORRELATION				
EFFECT	MS	df	F	p
GENOTYPE(G)	0.01	1	0.01	0.947
TEMP(T)	13.41	1	18.23	0.000*
LINE(L)	1.67	6	2.27	0.037*
T*G	0.97	1	1.32	0.251
L*T	1.14	6	1.55	0.162
ERROR	0.74	341		

Spectral

However with spectral data there was a significant Line effect and the Temperature Genotype interaction approached significance, while the Genotype main effect narrowly failed to reach significance (Appendix 5.10). On inspection of the spectral means (Fig. 5.11b) it is clear the two genotypes are antagonistic in effect with temperature, both settling on a period of approximately 24 h at the higher temperature. If one considers that the LEC30 data were omitted, the results from this line (period @ 18°C , 24.18 h; period @ 29°C ,23.7 h, Table 5.2) would have contributed further to this interaction (see also Fig. 5.3b).

Fig. 5.11a Mean locomotor activity periods
2-way interaction
F(1,341)=1.32; p<0.2512

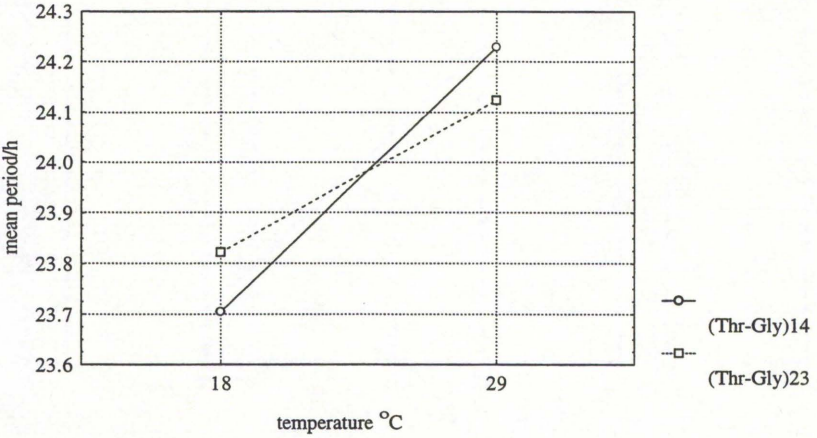


Fig. 5.11b Mean locomotor activity periods
2-way interaction
F(1,341)=2.91; p<0.0889

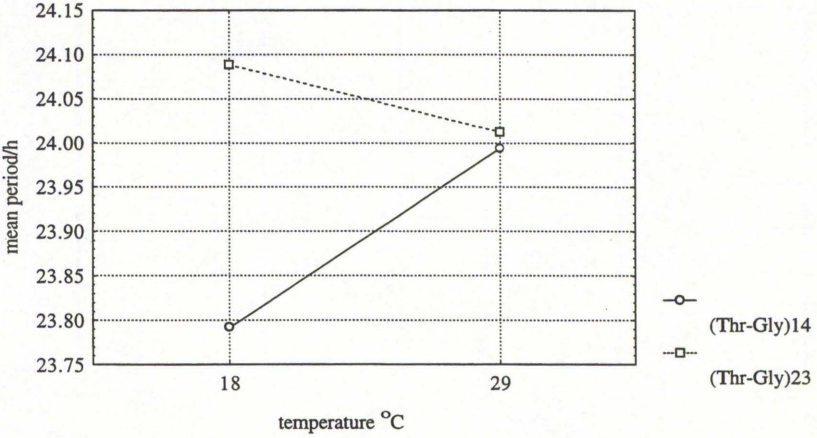


Figure 5.11a and b Means of locomotor activity period of $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ length variants at 18°C and 29°C for autotcorrelation and spectral respectively (note the near significant interaction coefficient for spectral data). These plots are different from Fig 5.4b since LECCE 30 (a $(Thr\text{-}Gly)_{23}$ allele) is excuded to balance the nested ANOVA.

In summary the spectral data for the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ length alleles reveal that the two variants may behave differently as can be seen in Fig. 5.11b. The Temperature Genotype interaction was not significant. In both graphs (Fig. 5.11a and b) the longer $(Thr\text{-}Gly)_{23}$ allele appears to be slightly better able to compensate for temperature than the shorter $(Thr\text{-}Gly)_{14}$ variants.

5.3.11 Analysis of northern and southern populations.

From the analyses which have examined the effect of lines within genotypes there appears to be significant heterogeneity (Tables 5.7, 5.8, 5.9). When examining the population/line data on Tables 5.1 and 5.2 it was noticeable that lines originally from southern Europe (S) and isolength lines isolated from northern Europe (N) differed in their temperature compensation ability. For example, the 3 lines carrying a $(Thr\text{-}Gly)_{20}$ allele that showed significantly different periods (spectral) at 18°C and 29°C (CAS10, LEC6 and LEC3) all originated from southern Europe. Also the two lines of $(Thr\text{-}Gly)_{17}$ length alleles that showed a poor temperature compensation (spectral) between temperatures were both from northern Europe, (CON17A and NW6). The lines were therefore divided into northern and southern populations; lines from Leiden (LE), North Wooton (NW), Cognac (CO) and Conselve (CON) were considered to be northern, spanning a latitudinal range of 52.16-45.23°, and lines emanating from Lecce (LEC), Pietrastornina (PI), Rethimnon (RET) and Casablanca (CAS) were regarded as southern with a latitudinal range of 40.35-33.66°. Both groups covered approximately 6-7° of latitude. A three way ANOVA was carried out for alleles carrying 14, 17, 20 and 23 Thr-Gly repeats, with Genotype, Temperature and North/South (N/S) as the main effects. The other length variants not included in this analysis represented either northern or southern groupings exclusively.

Autocorrelation

With autocorrelation, there was a significant Temperature (T) and Genotype, Northern/Southern interaction effect (G*N/S) , see Table 5.11. This suggests that there is a significant difference for period length between genotypes from different origins.

When comparing the isolength variants of the same population between temperatures, see Table 5.12, northern $(Thr\text{-}Gly)_{14}$ length variants, northern and southern

$(Thr\text{-}Gly)_{17}$ length variants and northern $(Thr\text{-}Gly)_{20}$ length variants had significantly different circadian periods at 18°C and 29°C.

Table 5.11 3-Way ANOVA for northern and southern populations of $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length alleles (*, $p<0.05$)

AUTOCORRELATION				
EFFECT	MS	df	F	p
GENOTYPE(G)	0.51	3	0.61	0.610
TEMP(T)	14.91	1	17.84	0.000*
N/S	1.22	1	1.45	0.228
T*G	1.22	3	1.45	0.225
G*N/S	4.40	3	5.26	0.001*
T*N/S	1.93	1	2.31	0.128
T*G*N/S	0.65	3	0.77	0.510
ERROR	0.84	1251		

In summary the $(Thr\text{-}Gly)_{17}$ length alleles had significantly different periods between the two temperatures whether they came from the north or the south of Europe. The northern $(Thr\text{-}Gly)_{20}$ variants were also significantly different between temperatures but the difference was smaller than that shown by the northern $(Thr\text{-}Gly)_{17}$ allele. The southern $(Thr\text{-}Gly)_{20}$ variants were not significantly different for period length between the two temperatures. These differences are illustrated in Fig. 5.12a and b, where the G*N/S interaction can clearly be seen with the $(Thr\text{-}Gly)_{17}$ allele giving different periods at different temperatures irrespective, of whether it is from northern or southern origins. As the three way interaction is not significant, the only other feature worth mentioning is that the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{17}$ alleles are slightly less well compensated in northern populations than the longer variants. The $(Thr\text{-}Gly)_{17}$ variant also reveals this feature in southern populations.

Spectral

With spectral periods a marginally significant Genotype effect was observed, but most critically both a Genotype Northern/Southern interaction (G*N/S) and a Temperature Genotype Northern/Southern interaction (T*G*N/S) showed significance, see Table 5.13. The Temperature Genotype interaction (T*G) just failed to reach significance. These

Fig. 5.12a, b. Mean locomotor activity periods
3-way interaction, Autocorrelation
F(3,1251)=.77; p<.5098

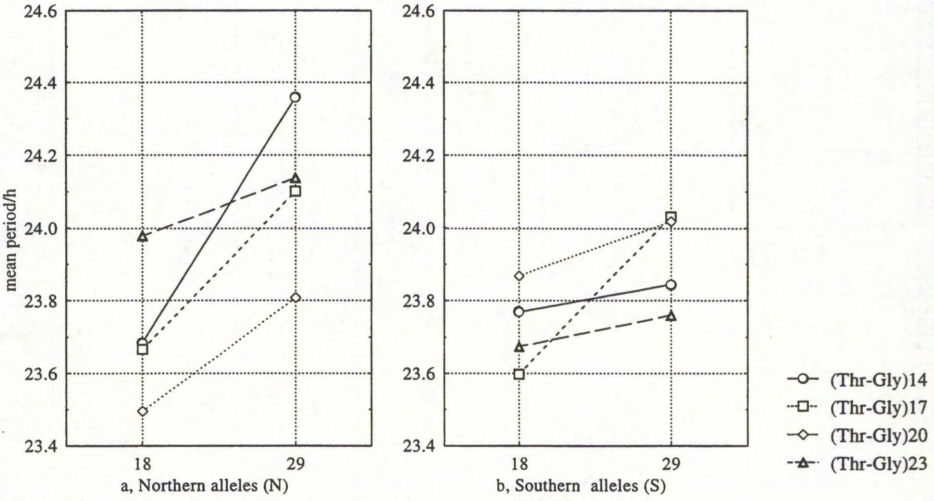


Figure 5.12a and b. Mean period length of northern and southern length alleles with respect to temperature for autocorrelation data.

Fig. 5.13a, b. Mean locomotor activity periods
, Spectral
F(3,1251)=3.84; p<0.0094

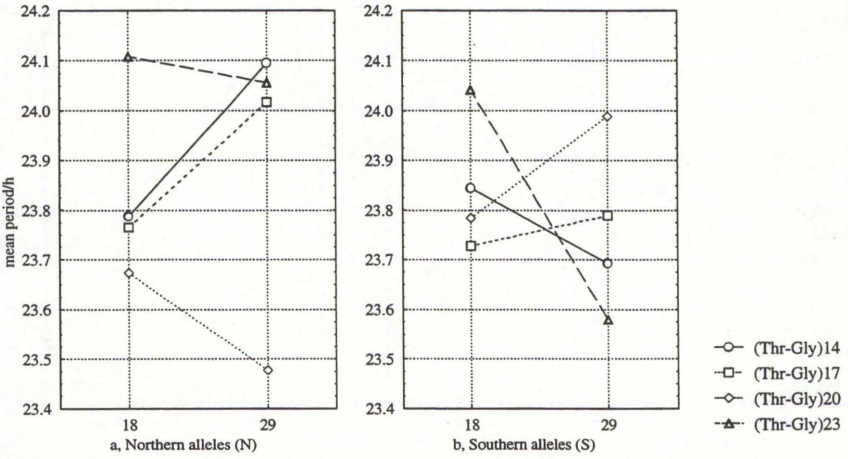


Figure 13 a and b. Mean period length of Northern and Southern length alleles with respect to temperature for Spectral data.

results together suggest that northern Thr-Gly length genotypes and southern Thr-Gly length genotypes are behaving differently at the two temperatures.

Table 5.12 Autocorrelation, Planned comparisons between alleles of the same origin at different temperatures (*, <0.05)

TEMP /^{o}C	Thr-Gly LENGTH	N/S	MEAN PERIOD/h	Difference between the means	MS	df	F	p
18	14	N	23.68	0.68	19.95	1	23.80	0.000*
29	14	N	24.36					
18	17	N	23.67	0.43	9.99	1	11.90	0.001*
29	17	N	24.10					
18	20	N	23.50	0.31	5.32	1	6.36	0.012*
29	20	N	23.81					
18	23	N	23.98	0.16	0.81	1	0.96	0.330
29	23	N	24.14					
18	14	S	23.77	0.07	0.04	1	0.05	0.830
29	14	S	23.84					
18	17	S	23.60	0.42	13.90	1	16.70	0.000*
29	17	S	24.02					
18	20	S	23.87	0.15	0.89	1	1.07	0.330
29	20	S	24.02					
18	23	S	23.68	0.12	0.08	1	0.09	0.760
29	23	S	23.76					

Table 5.13 3-WAY ANOVA for northern and southern populations of $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length alleles (*, p<0.05)

SPECTRAL EFFECT	MS	df	F	p
GENOTYPE(G)	1.66	3	2.61	0.050(*)
TEMP(T)	0.01	1	0.01	0.942
N/S	0.77	1	1.22	0.271
T*G	1.54	3	2.43	0.064
G*N/S	4.71	3	7.41	0.000*
T*N/S	1.20	1	1.89	0.169
T*G*N/S	2.44	3	3.84	0.009*
ERROR	0.64	1251		

When isolength populations from the same origin were compared (Table 5.14 and Figures 5.13a and b) at 18^{o}C and 29^{o}C , both northern $(Thr\text{-}Gly)_{14}$ and northern (Thr-

$Gly)_{17}$ lines had significantly different periods. The southern $(Thr\text{-}Gly)_{17}$ variants were not significantly different, suggesting some adaptation of southern $(Thr\text{-}Gly)_{17}$ alleles. However, the autocorrelation data for this allele was significant (see Table 5.12). The $(Thr\text{-}Gly)_{20}$ alleles for both northern and southern populations showed no significant difference between 18°C and 29°C, again showing an efficient temperature compensation. Additionally both the northern and southern $(Thr\text{-}Gly)_{23}$ length alleles showed no significant difference between temperatures. Inspecting the difference between the means it is evident that the southern $(Thr\text{-}Gly)_{17}$ allele is better compensated than the $(Thr\text{-}Gly)_{20}$ alleles from either the north or the south of Europe, whereas the northern $(Thr\text{-}Gly)_{17}$ and northern $(Thr\text{-}Gly)_{20}$ alleles have similar absolute differences between their means, although the $(Thr\text{-}Gly)_{17}$ variants difference is significant.

Table 5.14 Spectral, Planned comparisons between alleles of the same origin at different temperatures.(*, <0.05)

TEMP /°C	Thr-Gly-LENGTH	N/S	MEAN PERIOD/h	Difference between the means	MS	df	F	p
18	14	N	23.78	0.31	4.15	1	6.52	0.011*
29	14	N	24.09					
18	17	N	23.76	0.26	3.37	1	5.31	0.021*
29	17	N	24.02					
18	20	N	23.67	0.21	2.07	1	3.26	0.071
29	20	N	23.48					
18	23	N	24.11	0.05	0.08	1	0.13	0.717
29	23	N	24.06					
18	14	S	23.84	0.15	0.17	1	0.26	0.608
29	14	S	23.69					
18	17	S	23.76	0.03	0.28	1	0.44	0.507
29	17	S	23.79					
18	20	S	23.67	0.32	1.66	1	2.61	0.106
29	20	S	23.99					
18	23	S	24.11	0.47	2.28	1	3.60	0.058
29	23	S	23.58					

In summary, the northern $(Thr\text{-}Gly)_{17}$ length variants were less able to compensate for period length than the southern $(Thr\text{-}Gly)_{17}$ allele, whereas both the northern and southern $(Thr\text{-}Gly)_{20}$ length alleles were generally able to maintain their period length across temperature although there is some evidence from autocorrelation that the northern $(Thr\text{-}Gly)_{20}$ alleles were also slightly less efficient in temperature compensation than their

southern neighbours. These observations are illustrated in Figs. 5.13a and b where it is also noticeable that the southern $(Thr\text{-}Gly)_{23}$ is less able to maintain its period between the two temperatures than the northern $(Thr\text{-}Gly)_{23}$ allele.

5.3.12 Correlations between latitude and period length for Thr-Gly variants at different temperatures

As the origin of the different alleles appears to be important, correlations with latitudes were performed for individual fly periods for each genotype at each temperature for the different populations.

Autocorrelation

Table 5.15 reveals a significant correlation (Pearson, *r*) between period length and latitude was observed for the $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{21}$ and $(Thr\text{-}Gly)_{23}$ length variants at 29°C. No significant correlations were seen for period length at 18°C for any length variant see Table 5.15. When repeated using the Spearman rank order method (r_s, Appendix 5.11) almost identical results were obtained. Again no significant correlations were observed for period length and latitude at 18°C.

Table 5.15 Autocorrelation, Pearson product moment correlations for period length of Thr-Gly variants at 18°C and 29°C with latitude(*, p<0.05)

	18°C			29°C		
Thr-Gly Variant	*r*	p	df	*r*	p	df
$(Thr\text{-}Gly)_{14}$	-0.24	0.814	101	0.03	0.004*	99
$(Thr\text{-}Gly)_{17}$	0.04	0.520	264	0.06	0.325	244
$(Thr\text{-}Gly)_{20}$	-0.13	0.063	195	-0.12	0.111	181
$(Thr\text{-}Gly)_{21}$	-0.06	0.605	67	-0.32	0.011*	62
$(Thr\text{-}Gly)_{23}$	0.00	0.292	88	0.29	0.010*	79

Spectral

Even more convincing are the correlations (Pearson, *r*) observed using spectral data. The circadian periods of all length genotypes excluding the $(Thr\text{-}Gly)_{21}$ length allele were highly correlated with latitude at 29°C, see Tables 5.16 and Figs. 5.14a, b ,c and d. The $(Thr\text{-}Gly)_{17}$ allele is also highly correlated with latitude at 18°C, see Fig. 5.15a. It is interesting to note that the two most common European alleles, $(Thr\text{-}Gly)_{17}$ and (Thr-

Figure 5.14a
mean period = 20.392 + .08163 * latitude
Correlation: r = .21829
$(Thr\text{-}Gly)_{14}$ at 29 C

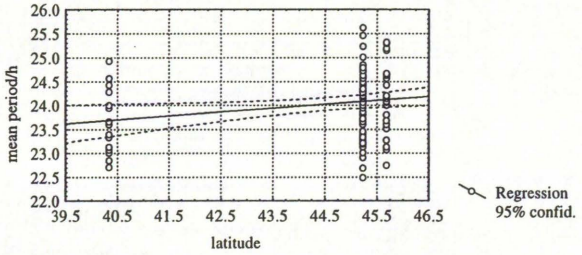


Figure 5.14b
mean period = 22.546 + .03220 * latitude
Correlation: r = .18634
$(Thr\text{-}Gly)_{17}$ at 29 C

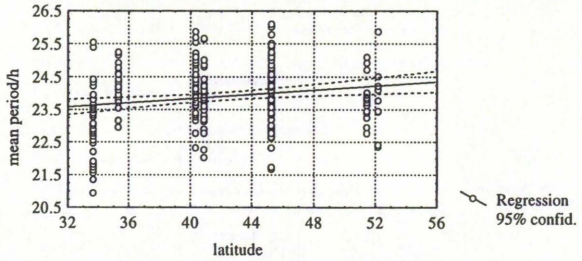


Figure 5.14c
mean period = 24.796 - .0258 * latitude
Correlation: r = -.1559
$(Thr\text{-}Gly)_{20}$ at 29 C

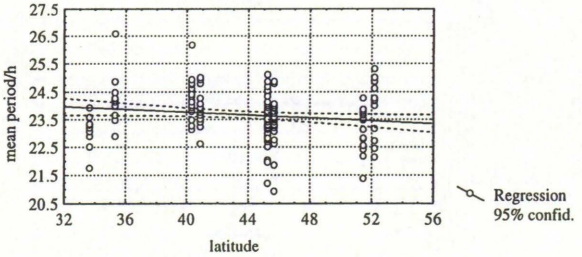
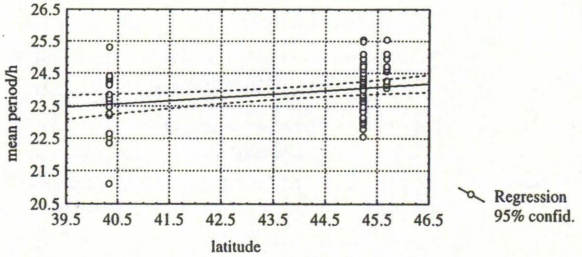


Figure 5.14d
mean period = 19.346 + .10422 * latitude
Correlation: r = .29346
$(Thr\text{-}Gly)_{23}$ at 29 C



Figures 5.14a, b, c and d Correlations of spectrally derived free-running locomotor activity period of individual Thr-Gly variant flies at 29 °C against latitude of origin (Pearsons *r*)

Fig 5.15a

mean period = 22.807 + .02282 * latitude

Correlation: r = 0.16575, p = 0.007*

(Thr-Gly)17 at 18 °C

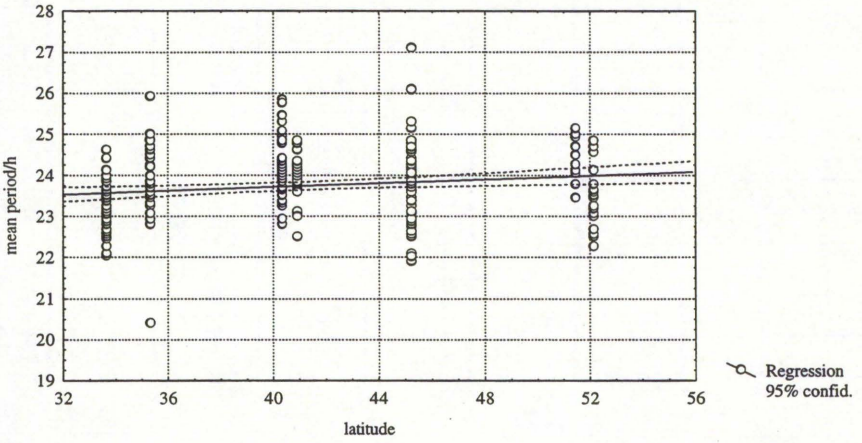


Fig. 5.15b

mean period = 23.788 - .0014 * latitude

Correlation: r = -0.0144, p = 0.841

(Thr-Gly)20 at 18 °C

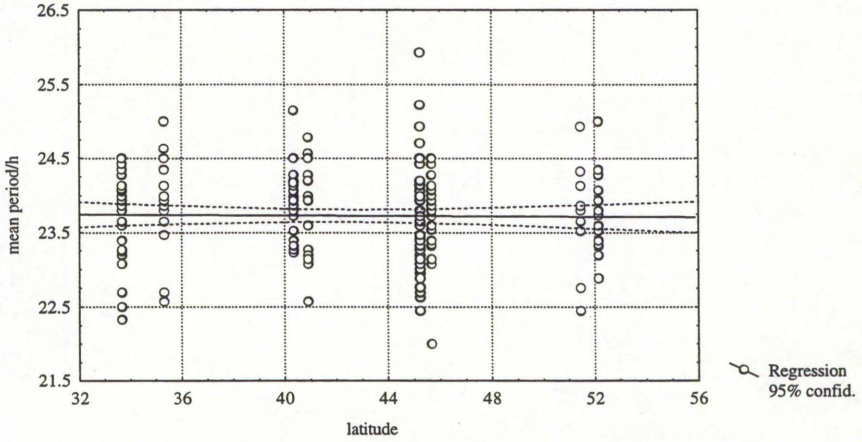


Figure 5.15a and b.Correlations of spectrally derived free-running locomotor activity periods of individual Thr-Gly variant flies at 18°C against the latitude of origin(Pearson's *r*)

$Gly)_{20}$, differ in the direction of their correlation with latitude. Whereas the former allele increases it's period length the further north its origin, the $(Thr\text{-}Gly)_{20}$ allele decreases its circadian period with an increase in latitude.

Table 5.16 Spectral, Pearson product moment correlations for period length of Thr-Gly variants at 18°C and 29°C (*, p<0.05)

	18°C			29°C		
Thr-Gly Variant	*r*	p	df	*r*	p	df
$(Thr\text{-}Gly)_{14}$	-0.01	0.907	101	0.22	0.028*	99
$(Thr\text{-}Gly)_{17}$	0.17	0.007*	264	0.19	0.003*	244
$(Thr\text{-}Gly)_{20}$	-0.01	0.841	195	-0.16	0.035*	181
$(Thr\text{-}Gly)_{21}$	-0.17	0.159	67	-0.06	0.645	62
$(Thr\text{-}Gly)_{23}$	0.06	0.610	89	0.29	0.008*	99

From observing the regression line of correlation for the $(Thr\text{-}Gly)_{20}$ length variant (Fig. 5.14c) at 29°C, it appears the southern alleles have a period closer to a 24 h rhythm than the northern alleles.

After using Spearman rank order correlations (r_s)on the same data at 29°C (Appendix 5.12) the identical highly significant correlations were seen with $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length alleles, but the $(Thr\text{-}Gly)_{14}$ length variant narrowly missing significance. No significant correlations were found for period length at 18°C apart from the $(Thr\text{-}Gly)_{17}$. However, it is interesting to note the direction of regression for the $(Thr\text{-}Gly)_{20}$ length allele is again opposite to that of the $(Thr\text{-}Gly)_{17}$ length allele at 18°C (Fig. 5.15b).

Correlations were also performed for the absolute differences between mean periods at different temperatures for lines within genotype (data in Tables 5.1 and 5.2) and latitude. No significant relationships were seen for any of the genotypes tested for either autocorrelation or spectral data. This is not surprising given that the data are based on lines, not individuals, and therefore very few data points are available for the correlation.

5.4 DISCUSSION

SUMMARY OF RESULTS

In summary this extensive behavioural and tortuous statistical analysis of the Thr-Gly length variants has revealed that they differ in their temperature compensation between 18°C and 29°C. The global analysis (section 5.3.1, Fig. 5.1 and 5.2) showed that the $(Thr\text{-}Gly)_{20}$ allele was the most efficient with respect to temperature compensation. This was further supported by the unnested analyses between the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ variants where a clear genotype effect was observed with spectral data (Appendix 5.5). When the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ alleles were compared in the same way the shorter allele appeared to have greater sensitivity to temperature (section 5.3.3, Fig. 5.4a, b).

When examining individual line data it was also obvious that the shorter length variants were more susceptible to period length differences with temperature changes (Table 5.1 and 5.2). Pooling lines within genotype, revealed that there are line differences within genotype but generally the longer length variants, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$, are better temperature compensated than the shorter $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{17}$ variants (Table 5.7 and 5.8).

After nesting lines within genotype for the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles, a weak genotype effect was observed (spectral) suggesting that the $(Thr\text{-}Gly)_{17}$ allele was slightly less efficient in temperature compensation than the $(Thr\text{-}Gly)_{20}$ (Appendix 5.8 and 5.9). When repeated with the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$, the longer allele again appeared to be better temperature compensated than the shorter allele (Section 5.3.10).

Analyses between northern and southern European lines revealed that the southern $(Thr\text{-}Gly)_{17}$ allele was less sensitive to temperature than its northern isolength allele and the southern $(Thr\text{-}Gly)_{20}$ variant (spectral, Fig. 5.13a and b), which in turn was less sensitive than its northern $(Thr\text{-}Gly)_{20}$ isolength allele. Autocorrelation data showed that both the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ differences in period length were quite similar across temperature for northern or southern origins

When using line means a significant correlation between temperature compensation and Thr-Gly length is observed for the 14, 17, 20 and 23 allelic series. This falls dramatically when the rarer alleles are included in the analysis (see Figs. 5.8a and 5.8b). This could be due to the fact that either fewer individuals are available for the rarer alleles,

giving a poorer correlation for simple stochastic reasons, or it could be a genuine effect. Arguing against the stochastic explanation is the observation that there are three $(Thr\text{-}Gly)_{21}$ alleles against only four of the more common $(Thr\text{-}Gly)_{14}$ length variant. Nevertheless in Fig. 5.8a and particularly 5.8b we still observe that the $(Thr\text{-}Gly)_{21}$ temperature compensation falls well off the regression line. If the $(Thr\text{-}Gly)_3$ stepwise increase between the common length alleles is important for temperature compensation, then it is of considerable interest that structural studies and molecular dynamics simulations have revealed that a $(Thr\text{-}Gly)_3$ peptide appears to form a stable type II or type III β-turn and this may represent the conformational monomer of the Thr-Gly repeat (Castiglione-Morelli *et al.* 1995).

Thr-Gly SELECTION ?

If the rarer alleles , $(Thr\text{-}Gly)_{17\text{-}2}$, $(Thr\text{-}Gly)_{21}$ and $(Thr\text{-}Gly)_{24}$ are less temperature compensated (Figs. 5.8a and 5.8b) this phenotypic difference could be relevant to the flies Darwinian fitness and therefore be a possible contributor to their extremely low frequencies across Europe as a whole. However, their low frequencies are much more likely to be due to the historical processes discussed in chapter 1. The relatively high mutation rate of the Thr-Gly repeat (Rosato *et al.*, ms. submitted) means that new variants will arise, but most will be eliminated by drift, or if neutral, they may remain at low frequencies. Novel length variations of this region have been found in Kenyan populations (C. Pasetto, pers. comm.) as would be expected in ancestral populations. Cann *et al.*(1987) and Vigilant *et al.*(1991) found that human mtDNA sequence variation was greater among Africans than non-Africans, suggesting an African origin for humans, as the occurrence and accumulation of mtDNA differences may be primarily a function of time. By analogy this suggests that the European *D. melanogaster* populations will have less variation due to founder effects during their early European migration. Phylogenetic analyses have revealed that the $(Thr\text{-}Gly)_{17}$ allele is derived from the $(Thr\text{-}Gly)_{20}$ (Rosato *et al.*, 1996), which in turn is derived from the ancestral $(Thr\text{-}Gly)_{23}$ allele. High frequencies of $(Thr\text{-}Gly)_{23}$ alleles have been found only in Africa where no $(Thr\text{-}Gly)_{17}$ allele has yet been found (C. Pasetto, pers. comm.) suggesting again that the $(Thr\text{-}Gly)_{17}$ variant is a relatively recent derivative of the $(Thr\text{-}Gly)_{20}$ variant and the mutational event responsible for its genesis is rare (Rosato *et al.*, 1996). Perhaps this event originated in northern Africa during the early migration of flies to Europe (circa 10, 000 years ago), and its adaptability to warmer climates gave it a selective

advantage, leading it to spread in southern Europe. Thus selection, partial admixture and historical effects probably all contributed to the cline.

FITNESS CHARACTERS

The differences seen between mean periods at different temperatures between genotypes are also quite illuminating if a 24 h circadian rhythm is assumed to be required for optimal fitness (Figs. 5.8a and 5.8b). Overall the $(Thr\text{-}Gly)_{17}$ variant has a precise 24h clock at the warmer temperature, whereas the $(Thr\text{-}Gly)_{20}$ variant although showing overall excellent temperature compensation fails to keep an exact 24 h period (Fig. 5.8a and 5.8b). This characteristic of the $(Thr\text{-}Gly)_{17}$ allele could be viewed as predisposing it to southern Europe where the overall temperature is higher and the annual temperature range is smaller and requires less response to thermal challenge. In the north the annual temperature range is greater, and perhaps this could explain why the $(Thr\text{-}Gly)_{20}$ alleles are found predominantly in northern Europe which has a harsher more variable climate. Inspection of the regression line for the $(Thr\text{-}Gly)_{17}$ length alleles activity periods against latitude at 29°C (Fig. 5.14a), also supports a selective interpretation for the cline, since southern populations of this allele have a period closer to 24 h than the northern populations at 29°C. Thus less physiological stress on these flies might result from the absence of resetting the clock by external zeitgeibers or light cues in the south. Pittendrigh and Minis (1972) showed with *D. melanogaster* that eukaryotic systems as oscillators performed most effectively when they were driven close to their natural circadian frequency, 24 h. A simple but compelling explanation for the European cline could be that it is the outcome of balancing selection between a well compensated clock (the $(Thr\text{-}Gly)_{20}$ allele clock) versus a precise 24 h clock (the $(Thr\text{-}Gly)_{17}$ allele clock) at higher temperatures. Another possibility might be that balancing selection could act at the level of heterozygote advantage. For example, if a heterozygous, $(Thr\text{-}Gly)_{17}/(Thr\text{-}Gly)_{20}$ female was phenotypically 'fitter' i.e. had a both a precise 24 h clock and 'perfect' temperature compensation compared to the two common variants, selection would favour both alleles. Although heterozygote advantage has yet to be investigated, some kind of balancing selection seems to be the best candidate for explaining the cline.

The observation of an opposite effect of latitude on activity periods at 29°C for the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles, the former increasing its period with an increase

in latitude and *vice versa* for the latter is interesting. Differences in photoperiod between northern and southern latitudes might be involved in this differentiation, but how is unclear. Periods increase and decrease with light intensity (Saunders 1982), and changes in circadian period often follow 'Aschoff's rule' where nocturnal animals increase their period with an increase in light intensity whereas, diurnal animals shorten their period (Pittendrigh, 1960). *Drosophila* are diurnal and might be expected to shorten their period with increasing light intensity. The $(Thr\text{-}Gly)_{17}$ allele follows this trend as it shortens its period the further south, where light intensity is greater. If the $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{17}$ carriers are differentially susceptible to light then differences in photoperiod might be important.

The capacity for temperature compensation is probably under the control of the whole Thr-Gly region not just the uninterrupted repeat, since a difference in period between the $(Thr\text{-}Gly)_{17}$ and the $(Thr\text{-}Gly)_{17\text{-}2}$ (a $(Thr\text{-}Gly)_{17}$ allele lacking the downstream $(Thr\text{-}Gly)_2$ repeat) was observed between temperatures (Fig. 5.8b). However, there was only one $(Thr\text{-}Gly)_{17\text{-}2}$ line found and analysed so this interpretation is preliminary. This deletion in *D. simulans* is very frequent and is in linkage disequilibrium with particular Thr-Gly length flanking alleles (Rosato *et al.*, 1994). However, in *D. melanogaster* it appears to be very rare and has only been found in two European individuals, one carrying a $(Thr\text{-}Gly)_{20}$ tract and the other the $(Thr\text{-}Gly)_{17}$ allele (Rosato *et al.*, 1996).

POPULATION ORIGIN

The story is complicated further by the fact that the temperature compensation of northern and southern length alleles differs within Thr-Gly genotype. For example, the southern $(Thr\text{-}Gly)_{17}$ alleles showed evidence for being better temperature compensated than their northern $(Thr\text{-}Gly)_{17}$ cousins (Table 14, Figs. 5.12a and b, 5.13a and b) whereas the $(Thr\text{-}Gly)_{20}$ length variant also showed some evidence for this, but only with autocorrelation (Fig. 5.12a and b, Table 5.12). The northern $(Thr\text{-}Gly)_{20}$ variant also showed a slightly smaller difference in mean period than the northern $(Thr\text{-}Gly)_{17}$ variant (Tables 5.12 and 5.14). The spectral data show that the southern $(Thr\text{-}Gly)_{17}$ alleles are better compensated than the southern $(Thr\text{-}Gly)_{20}$ variant, but this is not confirmed by the autocorrelation data (Tables 5.12 and 5.14). The $(Thr\text{-}Gly)_{23}$ allele however showed no significant difference for period length between temperatures when comparing northern and southern alleles. This result corresponds to the finding that this variant does not show a cline in Europe. However its slightly higher frequency in northern Europe (Costa *et al.*,

1992) could conceivably be related to its efficient temperature compensation which may provide some selective value in the north. The arbitrary nature of the northern and southern origin split of the European populations possibly masked any genuine effects of extreme spatial differentiation. This point will be examined further in chapter 7, when the Australian Thr-Gly length variants, taken from more spatially distinct origins, are analysed in DD cycles.

LINKAGE DISEQUILIBRIUM

It remains to be addressed whether any selection is acting on the Thr-Gly length polymorphism itself, or on closely linked variation. In the latter case the different length variants could be simply hitchhiking on closely linked variable sites that would themselves be under selection (Maynard-Smith and Haigh, 1974; Berry *et al.*, 1991; Begun and Aquadro, 1991, 1992). In *D. melanogaster* the repeat expands and contracts and this intraspecific change in length is not generally associated with amino acid changes in the flanking region (Rosato *et al.*, 1996). Only two flanking polymorphisms have been found, one involving a $(Thr\text{-}Gly)_{23}$ allele the other a $(Thr\text{-}Gly)_{17}$ variant (Rosato *et al.*, 1996; chapter 3, Fig. 3.4). Thus unlike interspecific differences in this region, where compensatory changes are observed in the flanking region to match the different repeat lengths (Peixoto *et al.*, 1992, 1993b; Nielsen *et al.*, 1994), in *D. melanogaster*, the length changes have no flanking amino acid changes associated with them. Chimeric transformants of *D. melanogaster* and *D. pseudoobscura*, where the repeat and flanking sequences are mixed and matched, show a drastic loss of temperature compensation. This could be the reason why differences in behaviour are associated with the *D. melanogaster* Thr-Gly variants.

An independent analysis of the Thr-Gly sequences has also suggested that the three major length variants, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ are a focus for weak selection (Rosato *et al.*, ms. submitted), and that selection may be operating on various haplotypes which share these Thr-Gly alleles. Basically the upstream (5') locus was defined by three silent nucleotide polymorphism at positions (nucleotide positions as in Citri *et al.*, 1987) 3166 (A/C), 3190 (G/A) and 3337 (C/T). The downstream locus was defined by 2 silent nucleotide polymorphisms at positions 5527 (C/T) and 5647 (A/T). The perfect Thr-Gly repeat is from position 5113 to 5234 based on the Oregon-R sequence (a $(Thr\text{-}Gly)_{20}$ allele, Citri *et al.*, 1987). Disequilibrium pattern analysis (Rosato *et al.*, ms. submitted)

suggested that selection was or is acting on loci of the AGC- $(Thr\text{-}Gly)_{20}$ -TA , CAC- $(Thr\text{-}Gly)_{17}$ -CA, length variant and CAT- $(Thr\text{-}Gly)_{23}$ -CT 3 locus haplotypes. By applying the method of Robinson *et al.* (1991) based on the comparison between the pairwise disequilibria existing in a three locus system, none of the three loci were able to be distinguished as the one under selection (Rosato *et al.*, ms. submitted) .

The differences in behaviour between *D. melanogaster* Thr-Gly length variants are unlikely to be caused by associated genetic background, as a large number of different *X* chromosomes from many different populations were studied on a randomised autosomal background. In addition the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles differ from each other by a single duplication/deletion event involving an 18 bp fragment which encodes $(Thr\text{-}Gly)_3$. Due to the repetitive nature of this region, the forward and reverse mutation rate of these two variants could be very high (Costa *et al.*, 1991). Linkage analysis has been used to infer an approximate mutation rate (μ) for the repeat of $10^{-5} < \mu < 4x10^{-5}$ per gamete per generation (Rosato *et al.*, ms. submitted). Consequently the Thr-Gly repeat does not have a mutation rate as high as some of the non-coding minisatellites (Jeffreys *et al.*, 1990, 1994), but it is several order of magnitude higher than the nucleotide substitution rate of 10^{-9} per generation (Kreitman, 1991). Therefore the idea that the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ variants are 'flipping' from one allelic state to the other at very high frequencies, thereby destroying any linkage disequilibrium between the repeat and flanking sites can be rejected, and thus it cannot be discounted that linked sites to the Thr-Gly are causing the behavioural changes (Rosato *et al.*, 1996; Costa *et al.*, 1991).

However no flanking haplotypes have been found to be in complete linkage disequilibrium with the different Thr-Gly variants (Rosato *et al.*, 1996). The sequence data for the behavioural lines tested (chapter 3) have not been investigated regarding these upstream and downstream haplotypes previously described, but it is unlikely that they would show any differences from the patterns examined by Rosato *et al.* (1996). The populations examined for linkage disequilibrium (Rosato *et al.*, ms. submitted), when separated into northern and southern populations, provide no specific three loci haplotypes for either origin that can explain the difference seen between the temperature compensation of spatially differentiated populations. This indicates that it is the Thr-Gly region itself under selection and not any changes in both the upstream and downstream sequences.

FURTHER EVIDENCE FOR A Thr-Gly MEDIATED TEMPERATURE COMPENSATION MECHANISM

Perhaps the strongest evidence that the differences seen in the different natural Thr-Gly variants for temperature compensation are real and are due to the repeat itself come from the work of my colleagues J. M Hennessy and H. Parkinson who have been deleting the Thr-Gly region of *per in vitro* and transforming these *per* constructs back into per^{01} transformants. All these constructs were internal deletions of an original 13.2 kb cloned *per* gene (Citri *et al.*, 1987), and so there is no linkage disequilibrium associated with the repeat region. Furthermore each construct is uniquely and randomly placed around the genome. Firstly a larger deletion than that of Yu *et al* (1987b) was made (H Parkinson pers. comm.). This deletion stretches over the postulated 'spacer' region of the protein (Costa *et al.*, 1991; Castiglione-Morelli *et al.*, 1995). Also transformant flies carrying either a single Thr-Gly repeat, $(Thr\text{-}Gly)_1$, a $(Thr\text{-}Gly)_{17}$ length allele, and a $(Thr\text{-}Gly)_{20}$ length allele were tested for temperature compensation. The original Δ(Thr-Gly) (Yu *et al.*, 1987b), which is $(Thr\text{-}Gly)_0$ and has a few flanking amino acids deleted was also examined, Table 5.17 shows the amino acids deleted for each construct, and all flies have the same amino acid sequence except for the deletions. Fig. 5.16 shows the mean locomotor activity periods for all constructs at 18°C, 25°C and 29°C, and the results shown strengthen the view that the length of the Thr-Gly region itself may be causing the temperature sensitivity of the behavioural phenotype.

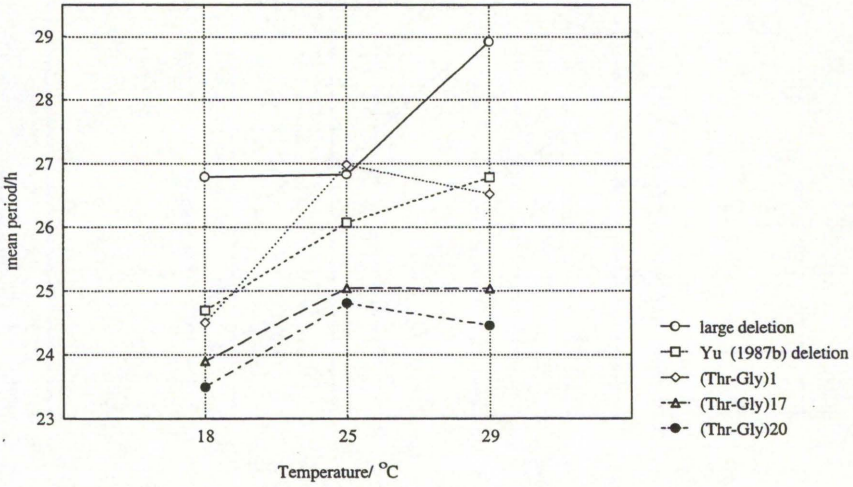
Table 5.17

CONSTRUCT	Amino acids deleted (Citri *et al.*, 1988)
large deletion	5102-5307
Δ (Thr-Gly) Yu *et al.*, (1987b)	5078-5218
$(Thr\text{-}Gly)_1$	5119-5233
$(Thr\text{-}Gly)_{17}$	5173-5191
$(Thr\text{-}Gly)_{20}$	as 13.2 kb construct (no deletion)

Fig. 5.16 shows that flies carrying a $(Thr\text{-}Gly)_{20}$ transformant appear extremely well compensated for locomotor activity rhythms , whereas the corresponding $(Thr\text{-}Gly)_{17}$ length transformants show a slightly more temperature sensitive phenotype. As mentioned

previously, when the Thr-Gly region plus some flanking sequences are deleted (Δ(Thr-Gly)) and then tranformed into per^{01} flies, locomotor activity periods lengthen with temperature (Ewer *et al.*, 1990 and Fig. 5.16). When per^{01} flies are tranformed with a *per* gene where only one Thr-Gly pair remains, their locomotor activity period increases between 18°C and 25°C followed by a shortening to approximately 26.5 hours at 29°C. (Fig. 5.16). Flies carrying the largest deletion of the Thr-Gly region have a lengthened period of 26.5h at 18°C and 25°C and this increases to 28.9h at 29°C (Fig. 5.16), indicating that this deletion is more deleterious than the former Yu *et al.* (1987b) Δ(Thr-Gly) deletion, perhaps because it removes the whole of the flexible region (Costa *et al.*, 1991; Castiglione-Morelli *et al.*, 1995). These results show a simple linear relationship between the Thr-Gly length and temperature compensation, and provide additional compelling evidence that the effect seen in natural variants, where length differs less dramatically, are indeed real.

Fig. 5.16 Mean periods of locomotor activity of the *in vitro* deleted Thr-Gly constructs (J. M. Hennessy and H Parkinson unpublished results)



MODELS OF TEMPERATURE COMPENSATION

How might the Thr-Gly domain mediate its effects on temperature compensation? As previously mentioned structural studies and molecular dynamic simulations have

revealed that a $(Thr\text{-}Gly)_3$ peptide forms a stable beta-turn under favourable conditions and that this structural motif may represent the conformational monomer. These peptides and polypeptides show very flexible conformations, especially at higher temperatures (Ishida *et al.* 1994; Castiglione-Morelli *et al.* 1995), and perhaps this relates in some way to the differences seen in temperature compensation for the Thr-Gly variants. Perhaps the flexible region acts like a spacer to maintain the separated PER globular domains in a correct orientation. Addition of energy to this system in the way of temperature, may in the absence of such flexibility, alter the orientation of the domains with respect to either themselves or the putative molecules that interact with PER (Huang *et al.*, 1993, 1995; Vosshall *et al.*, 1994; Wietz *et al.*, 1995). A longer repeat rather than a shorter one may be necessary to provide sufficient flexibility to the PER molecule at lower temperatures. Furthermore, the longer repeat at low temperatures may also provide a weaker elastic tension along the length of the amino acid chain, whereas a shorter repeat may be enough to provide the necessary flexibility at higher temperatures. Thus, a subtle balance between Thr-Gly length and any associated flexibility may guarantee a constant circadian phenotype at different temperatures.

The nuclear translocation of PER is likely to be regulated by the product of a recently isolated clock gene, *timeless* (*tim*) (Seghal *et al.*, 1994). In tim^{01} mutants loss of behavioural rhythms (Myers *et al.*, 1995) and loss of oscillating *per* mRNA levels (Seghal *et al.*, 1994) is accompanied by a failure of PER to undergo nuclear translocation (Vosshall *et al.*, 1994). These observations suggest that the *tim* gene regulates the timing of the PER negative feedback loop. Huang *et al.*(1993, 1995), proposed that the PER PAS domain mediates homotypic intermolecular interactions. More recently, TIM seems to have become the ideal candidate for this interaction. Furthermore PER PAS also undergoes an intramolecular interaction with the downstream C-domain (Huang *et al.*, 1995). This intramolecular interaction would result in a closed PER conformation and would compete with the PAS mediated intermolecular alliance, see Fig. 5.17. Because both sets of interactions share common components, their temperature coefficients would be similar and as a consequence, PER dimerisation should manifest temperature compensation. As the per^{LI} mutation results in an enhanced PAS C-domain interaction with an increase in temperature, the availability of PER for heterotypic dimerisation with TIM is reduced and consequently it's translocation into the nucleus is delayed (Huang *et al.*, 1993; 1995). The intramolecular interaction thus could contribute to nuclear entry because per^{LI} nuclear

translocation is temperature sensitive, and is correlated with the mutants effects on period length (Curtin *et al.*, 1995; Huang *et al.* 1995).

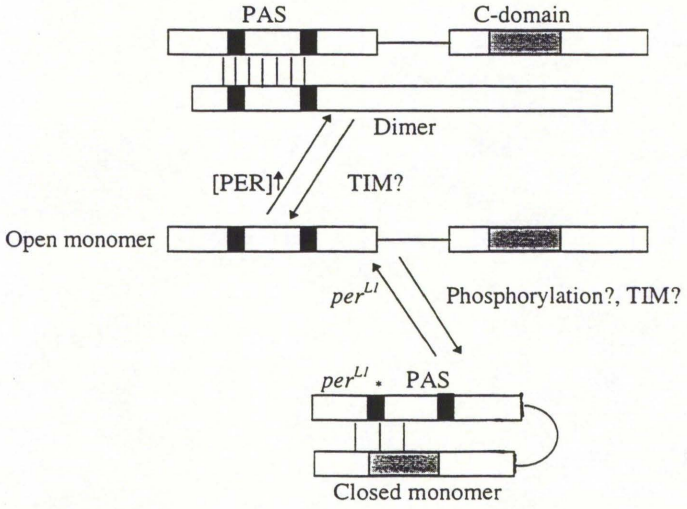


Figure 5.17 A hypothetical equilibrium between the PER closed monomer, and the dimer. The PAS domain, C-domain, and per^{L1} mutation are indicated and TIM is shown to dimerise with PER. (redrawn from Huang *et al.*, 1995)

Using the yeast two hybrid assay Gekakis *et al.*(1995) demonstrated that the interaction between TIM and PER^{L1} is characterised by a grossly abnormal temperature dependence. These results suggest that the temperature sensitive decrease in the affinity of PER^{L1} for TIM could account for the temperature sensitive delay of PER nuclear entry and the temperature sensitive lengthening of the circadian period in per^{L1} flies. These findings excluded the former proposal (Huang *et al.*, 1995), at least with respect to the PER-TIM interaction, that the per^{L1} mutation disrupts the association of PER with other proteins by means of a temperature dependent, enhanced intramolecular alliance between PAS and the C-domain. This is because PER^{L1} entirely lacking the C-domain, shows the same temperature sensitivity interactions with TIM as does PER^{L} with the C-domain (Gekakis *et al.*, 1995).

The Thr-Gly region could be involved structurally in both any intramolecular and intermolecular interactions. Perhaps the longer and therefore more flexible the region, the easier it is for these intra- and intermolecular interactions to occur

A direct intramolecular association between the PER PAS domain and the Thr-Gly domain is suggested by the following ongoing work in our laboratory. Hybrid *per* genes from *D. melanogaster* and *D. pseudoobscura*, which have different repeat lengths (Petersen *et al.*, 1988; Costa *et al.*, 1991), have been generated where the chimaeric junctions were designed to separate the repeat of one species from the flanking region of the other, thereby destroying the proposed coevolution between them (Peixoto *et al.*, 1993b, Nielsen *et al.*, 1994). A series of studies using this approach fully supported the idea of coevolution since some 'coevolution disrupted' transformants showed almost arrhythmic phenotypes while others showed a loss of temperature compensation (A.A. Peixoto *et al.*, ms. in preparation). These dramatic results show that the length of the repeat and its flanking region had coevolved to maintain the thermostability of the clock (A.A. Peixoto ms. in preparation, I Townson pers. comm.). Furthermore one of these hybrid constructs *mps5*, where repeats and flanking regions were mixed and matched gives a temperature sensitive per^{L1} like phenotype (A. A Peixoto ms. in preparation, I. Townson pers. comm.). If the Thr-Gly region physically interacts with the PAS domain then perhaps a per^{L1} phenotype can be generated by mutations in either PAS or the Thr-Gly regions. Preliminary results using the yeast two hybrid system suggest that this may be the case (I. Townson pers. comm.). It would be extremely interesting to carry out yeast two hybrid assays to see whether the Thr-Gly region interacts with TIM. Perhaps the entry of the TIM PER complex into the nucleus is gated by factors which include the Thr-Gly region.

The results from several different lines of enquiry suggest that the Thr-Gly region appears to be integral in temperature compensation. Furthermore the experiments reported in this chapter can be seen within the 'coevolution' framework outlined above, that each *D. melanogaster* Thr-Gly length variant does not have a 'compensating' amino acid linked to it, unlike the interspecific differences. This therefore leads to temperature sensitivities in circadian behaviour. The natural Thr-Gly length polymorphism, shaped by natural selection, in effect, 'fine tunes' the temperature compensation mechanism to particular thermal environments.

In conclusion the natural *per* alleles have therefore led us to propose a selective explanation for the spatial differentiation of the Thr-Gly length variation seen in Europe and North Africa.

CHAPTER 6

BEHAVIOURAL ANALYSIS OF TWO EUROPEAN POPULATIONS IN LIGHT/DARK CYCLES

6.1 INTRODUCTION

In the previous chapter a behavioural analysis of different European Thr-Gly length variants was described, and it was revealed that the Thr-Gly alleles differed in their temperature compensation between 18°C and 29°C. More specifically the $(Thr\text{-}Gly)_{20}$ length alleles, when individuals were pooled within genotypes, showed more efficient temperature compensation than the shorter $(Thr\text{-}Gly)_{17}$ length variant. In addition the southern $(Thr\text{-}Gly)_{17}$ length alleles showed some evidence of improved temperature compensation compared to their northern isolength neighbours, particularly with spectral analyses.

It is easy to understand the idea that the $(Thr\text{-}Gly)_{20}$ length variant is more abundant in the north due to its better overall temperature compensation compared to the $(Thr\text{-}Gly)_{17}$ allele. From Figs. 5.13 a and b, there is no evidence that southern $(Thr\text{-}Gly)_{17}$ are better compensated than the $(Thr\text{-}Gly)_{20}$ alleles. Overall, the $(Thr\text{-}Gly)_{17}$ does keep closer to a 24h period than the $(Thr\text{-}Gly)_{20}$ at higher temperatures (Figs. 5.8a and b), and I have argued that this might be a factor favouring the $(Thr\text{-}Gly)_{17}$ variant in warmer climates. However, even this observation is not supported by southern $(Thr\text{-}Gly)_{17}$ variants whose periods are no closer to 24h than the southern $(Thr\text{-}Gly)_{20}$ variants at any temperature. Perhaps the somewhat arbitrary division into northern and southern populations is not focused sharply enough on this point. There may be other reasons for selection of the $(Thr\text{-}Gly)_{17}$ variant at lower latitudes other than by the historical effects discussed in chapter 4 and 5. Perhaps the $(Thr\text{-}Gly)_{17}$ allele encodes an advantageous property which is not required in the cooler north, and which is not exhibited by the $(Thr\text{-}Gly)_{20}$ allele. Consequently I have studied a sample of *D. melanogaster* Thr-Gly length variant lines from Conselve and Lecce, in a LD 12:12 regime, in order to investigate any phenotypic differences in the distribution of locomotor activity at both 18°C and 29°C.

6.2 METHODS

Attached-X lines already genotyped in chapter 5 for Thr-Gly repeat length from Conselve and Lecce, both Italian populations, were used in this investigation, see Table 5.1. These two populations were used since they each represented almost the entire range of natural Thr-Gly length alleles found in Europe at that time and therefore could be considered as replicate populations. There were three $(Thr\text{-}Gly)_{14}$ lines (LEC14, CON 14A and B), five $(Thr\text{-}Gly)_{17}$ lines (LEC60, LEC12, CON 17A, B and C), four $(Thr\text{-}Gly)_{20}$ lines (LEC6, LEC3, CON20A and B), two $(Thr\text{-}Gly)_{21}$ (CON21A and B) and four $(Thr\text{-}Gly)_{23}$ lines (LEC79, LEC30, CON20B and B) analysed. Flies were raised in LD 12:12 at 25°C and then entrained for 24h at either 18°C or 29°C before locomotor activity was recorded for the following seven days. Lights on was at ZT0 (9am) and lights off at ZT12 (9pm).

Statistical analysis of the data was carried out using Minitab version 8 software and Statistica (Statsoft) for Windows. Initially the data recorded was 'wrapped', that is, a mean activity for each of 48, 30min bins across the seven days was calculated by superimposing the data for each of the seven days on top of each other, giving a representation of an average days mean activity for each fly. The genotype means were then calculated per bin. Following this the 48 bins were divided into 12, 2h (four bin) windows, starting from ZT6 (3pm) since this was the time that data recording was initiated, and a mean activity for each 2h window was calculated. Next, each of these values was converted into a proportion of the total activity and then arcsine transformed. However, these transformed proportions are no longer absolutely independent since they reflect a proportion of the total activity.

6.3 RESULTS

The *a priori* hypothesis was that flies carrying different Thr-Gly length genotypes would show differences in their activity distributions at different temperatures. As a guide, any effect that could be replicated in both populations would be of considerable interest and this might enhance any selective explanation of the European cline found for the two most common length variants the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$.

6.3.1 ANOVA for raw activity

A separate threeway ANOVA was carried out on Conselve and Lecce populations with Genotype (5 levels), Temperature (2 levels) and Window (12 levels) as the main effects, window 1 being activity between ZT6-8, window 2 between ZT8-10 and so on. Individual planned comparisons between $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ carriers and between $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ variants at both temperatures for each 2h window were also performed for the Conselve and Lecce populations.

Conselve

All effects were significant , except the T*G*W, which just failed to reach significance (Table 6.1). Figure 6.1 also shows the distribution of activity for each genotype. For both temperatures it can be seen that the $(Thr\text{-}Gly)_{20}$ carrier is considerably more active than the $(Thr\text{-}Gly)_{17}$ allele during the day.

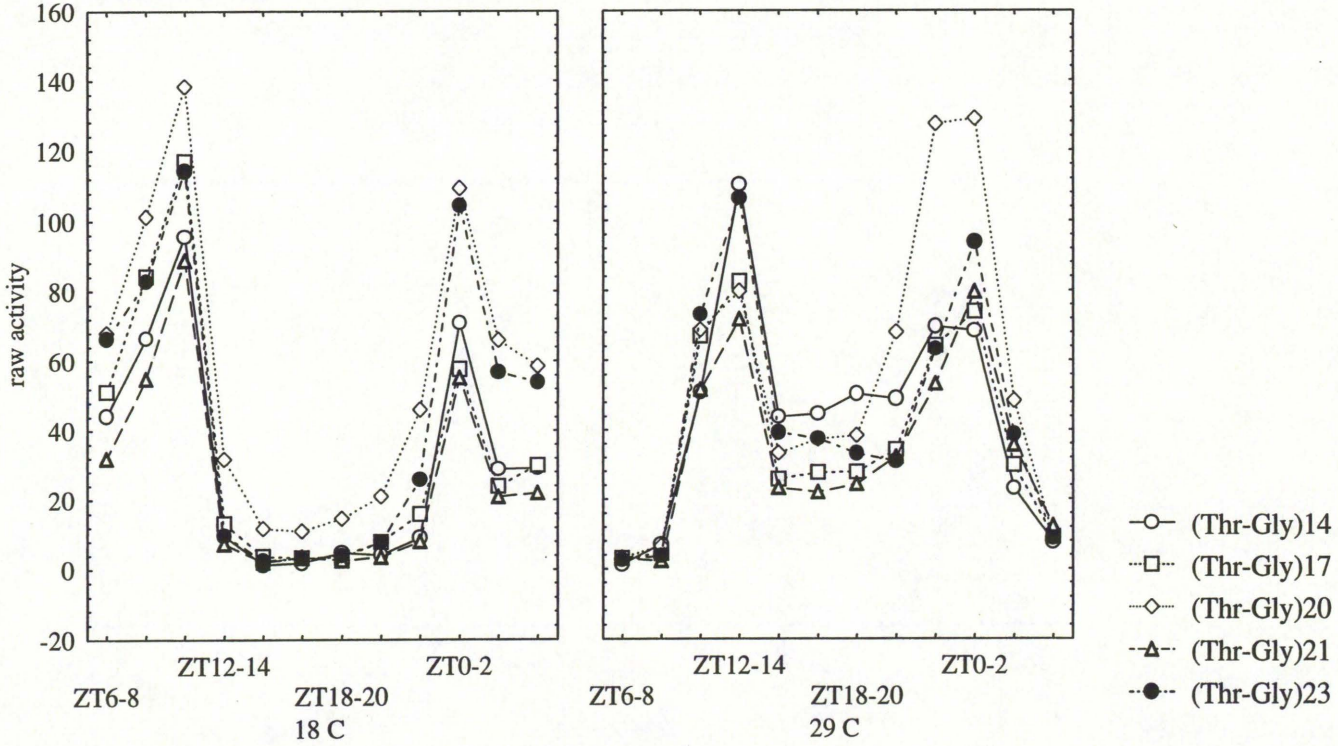
TABLE 6.1 ANOVA of raw activity (*, p<0.05)

CONSELVE				
EFFECT	MS	df	F	p
GENOTYPE(G)	56868.80	4	33.82	0.000*
TEMP(T)	15683.70	1	9.33	0.002*
WINDOW(W)	136294.40	11	81.05	0.000*
T*G	5521.00	4	3.28	0.011*
W*T	127291.10	11	75.70	0.000*
T*G*W	2268.20	44	1.35	0.063
ERROR		2952		

Fig. 6.1 Raw activity means of the Thr-Gly variants

3-way interaction, Conselve

$F(44,2952)=1.35$; $p<0.0628$



Planned comparisons were carried out between the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length variants and also between the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ carriers (Appendix 6.1 and 2). Significant differences were seen for bins ZT22-6 at 18°C and ZT20-2 at 29°C when comparing the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ variants. When testing the $(Thr\text{-}Gly)_{14}$ with the $(Thr\text{-}Gly)_{23}$ length variant only bins ZT0-6 were significantly different at 18°C. No notable differences were observed at 29°C, although ZT0-2 did narrowly miss significance.

For clarity, Fig. 6.2 shows the activity profiles of the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ only, with the asterisks representing significant differences in raw activity between them when planned comparisons were carried out. Generally the Conselve $(Thr\text{-}Gly)_{20}$ variants are more active than the $(Thr\text{-}Gly)_{17}$ at 18°C and 29°C, but only specifically from ZT18 to ZT0 (Fig. 6.2)

Lecce

When performing the same analyses with the Lecce lines an identical set of significant effects was observed compared to Conselve (Table 6.2).

TABLE 6.2 ANOVA of raw activity (*, $p<0.05$)

LECCE				
EFFECT	MS	df	F	p
GENOTYPE(G)	7052.70	3	6.11	0.000*
TEMP(T)	36738.00	1	31.81	0.000*
WINDOW(W)	111257.20	11	96.34	0.000*
T*G	5536.00	3	4.79	0.002*
W*T	39533.90	11	34.23	0.000*
T*G*W	1509.70	33	1.31	0.114
ERROR	1154.887	2316		

Planned comparisons between $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ length variants respectively (Appendix 6.3 and 6.4) also revealed few differences. Most significantly the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ variants differed in raw activity for the ZT0-2 window at 29°C, but the direction of this difference was opposite to that found for Conselve, whereas, the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ alleles differed from each other in the same window at 18°C. Figure 6.3 shows the raw activity profiles of all the Lecce Thr-Gly

Fig. 6.2 Raw activity means of $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$

Conselve (* denotes $p<0.05$)

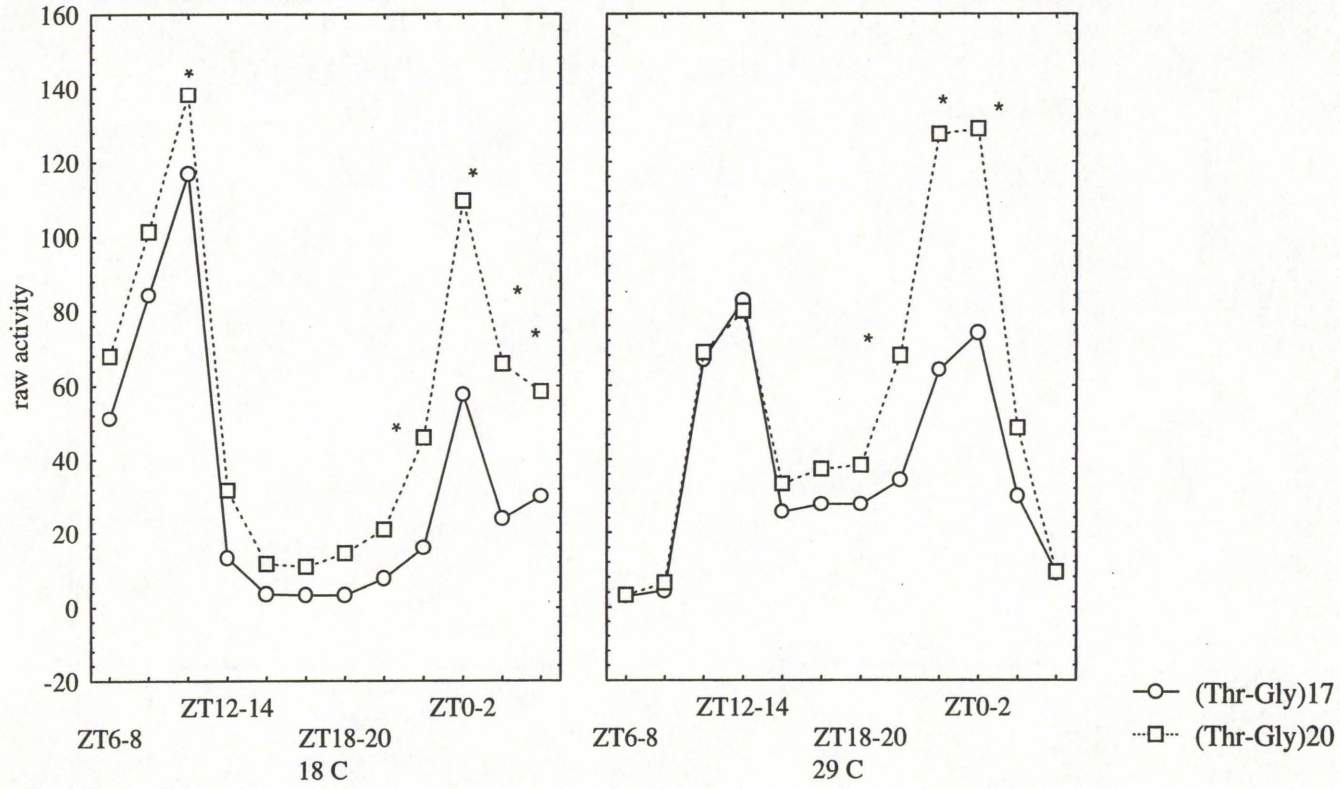
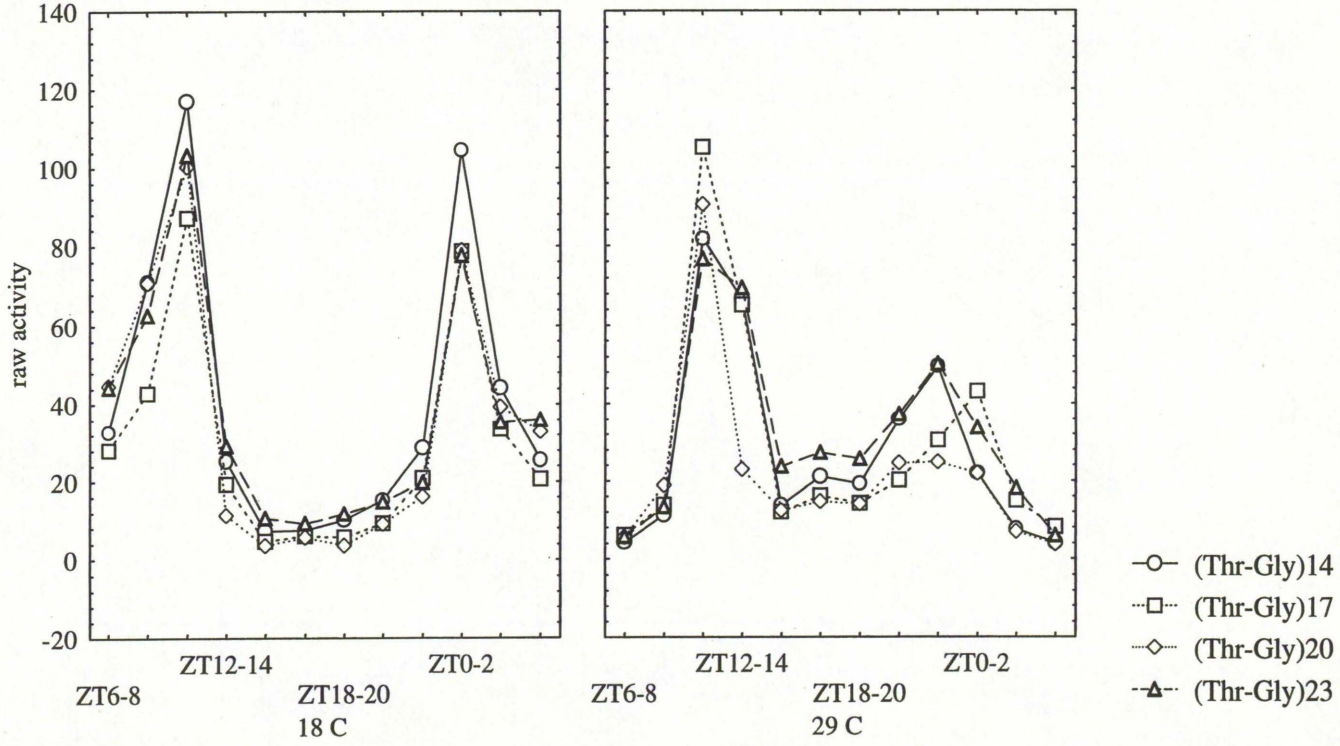


Fig. 6.3 Raw activity means of the Thr-Gly variants

3-way interaction, Lecce

F(33,2316)=1.31; p<0.1136



genotypes, while Fig. 6.4 illustrates only the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ variants raw activity profiles and any significant differences between them.

Lecce $(Thr\text{-}Gly)_{20}$ variants were therefore generally more active than the $(Thr\text{-}Gly)_{17}$ at 18°C between ZT2-and ZT10, but not at 29°C (Fig. 6.4).

In summary, these analyses revealed differences between the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ variants. However, the only consistent effect between populations was that the $(Thr\text{-}Gly)_{20}$ carriers were slightly more active at 18°C than the $(Thr\text{-}Gly)_{17}$ allele.

6.3.2 Correlations between raw activity and Thr-Gly length at the two temperatures

Correlations between Thr-Gly length and activity were performed for each 2h window for both the 18°C and 29°C data.

Conselve and Lecce

Four 2h windows were significantly correlated with Thr-Gly length at 18°C (Appendix 6.5), these windows covered ZT22-6, Figs. 6.5a to d shows these correlations as plots of the raw activity per time window for each Thr-Gly length variant. Two significant positive correlations were observed at 29°C for ZT0-2 and ZT2-4, Figs. 6.6a and b, suggesting higher activity during these time periods by the longer length alleles. A single significant correlation was observed with the Lecce lines, for ZT2-4 at 29°C (Appendix 6.6).

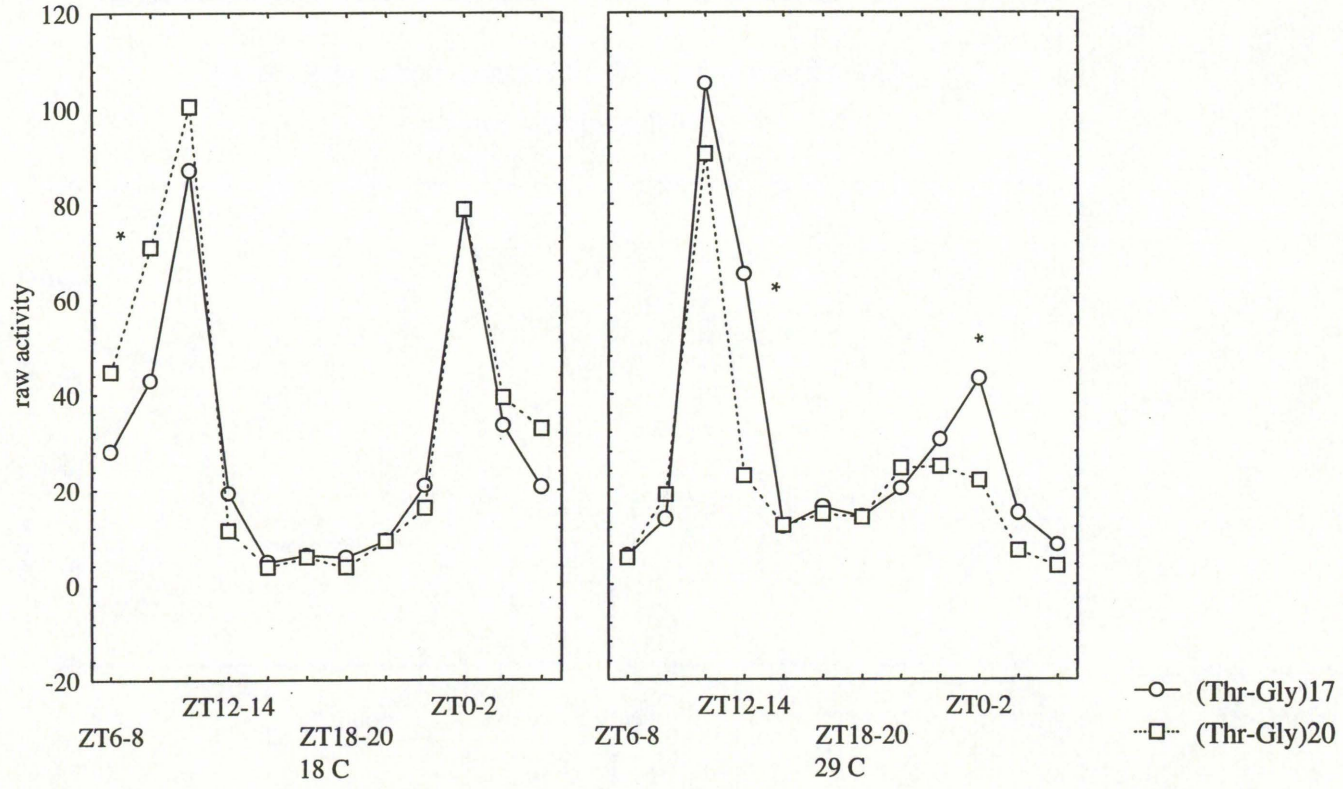
6.3.3 ANOVA of the Thr-Gly variants (transformed proportions)

An ANOVA was carried out on all genotypes for both populations using the transformed proportions of the raw activity data. These proportions of activity reflect flies' division of their energetic output across the day.

Conselve

All effects except T*G were significant, including a significant interaction effect for the T*G*W suggesting that the different Thr-Gly length variants were behaving differently with respect to temperature within different 2h windows (Table 6.3).

Fig. 6.4 Raw activity means of the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ Lecce (*denotes $p<0.05$)



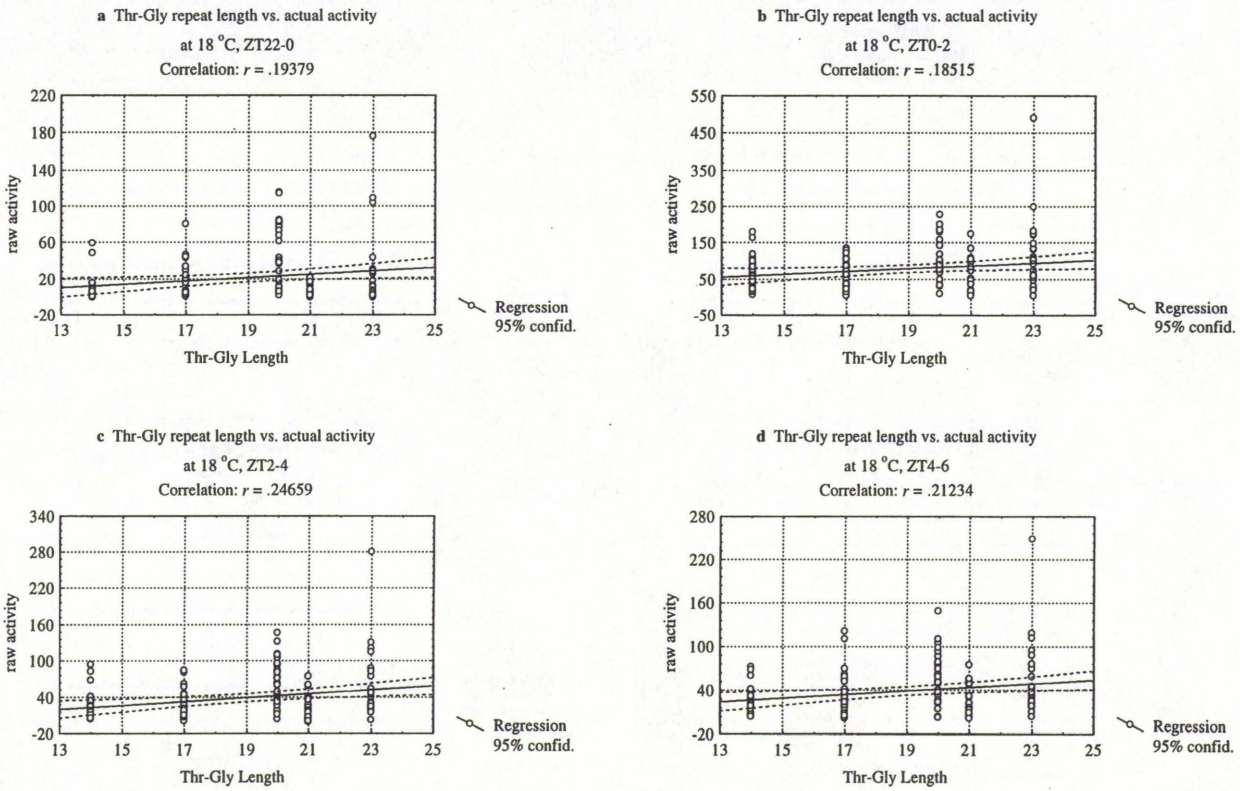


Figure 6.5a to d Correlations of the raw activity with Thr-Gly length for particular 2h windows at 18 °C, CONSELVE

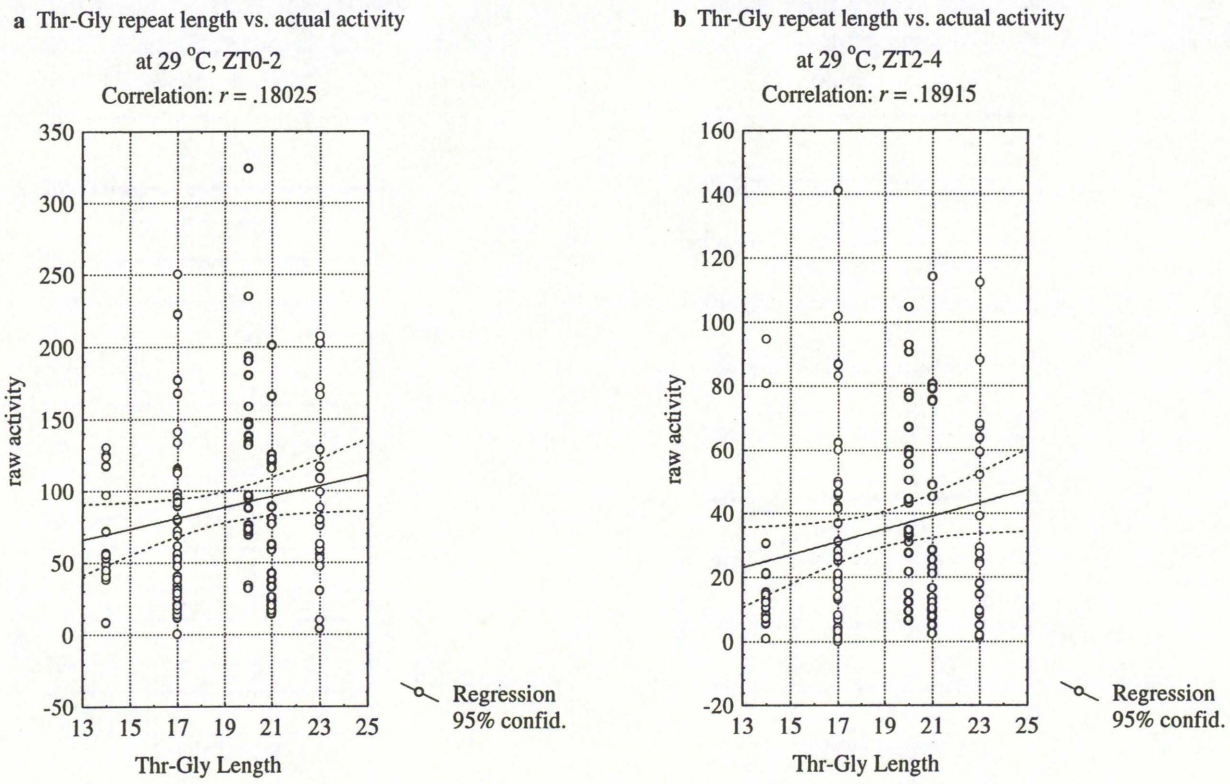


Figure 6.6a and b.Correlations of raw activity with Thr-Gly length for particular 2h windows at 29 °C, CONSELVE

TABLE 6.3 ANOVA of the Thr-Gly alleles (transformed proportions, *, $p<0.05$)

CONSELVE				
EFFECT	MS	df	F	p
GENOTYPE(G)	0.21	4	2.85	0.023*
TEMP(T)	0.58	1	7.76	0.005*
WINDOW(W)	2.52	11	335.17	0.000*
T*G	0.01	4	1.89	0.109
W*T	2.64	11	351.83	0.000*
T*G*W	0.03	44	3.47	0.000*
ERROR	0.075	2952		

Fig 6.7 reveals that although similar patterns of relative activity distribution can be observed, slight differences between Thr-Gly length variants can also be seen, for example the $(Thr\text{-}Gly)_{20}$ length variant appears to have a higher proportion of relative activity during the night than the other Thr-Gly length alleles at 18°C. Planned comparisons using the two most common alleles and the shortest and longest length alleles tested confirmed these observations (Appendix 6.7 and 6.8).

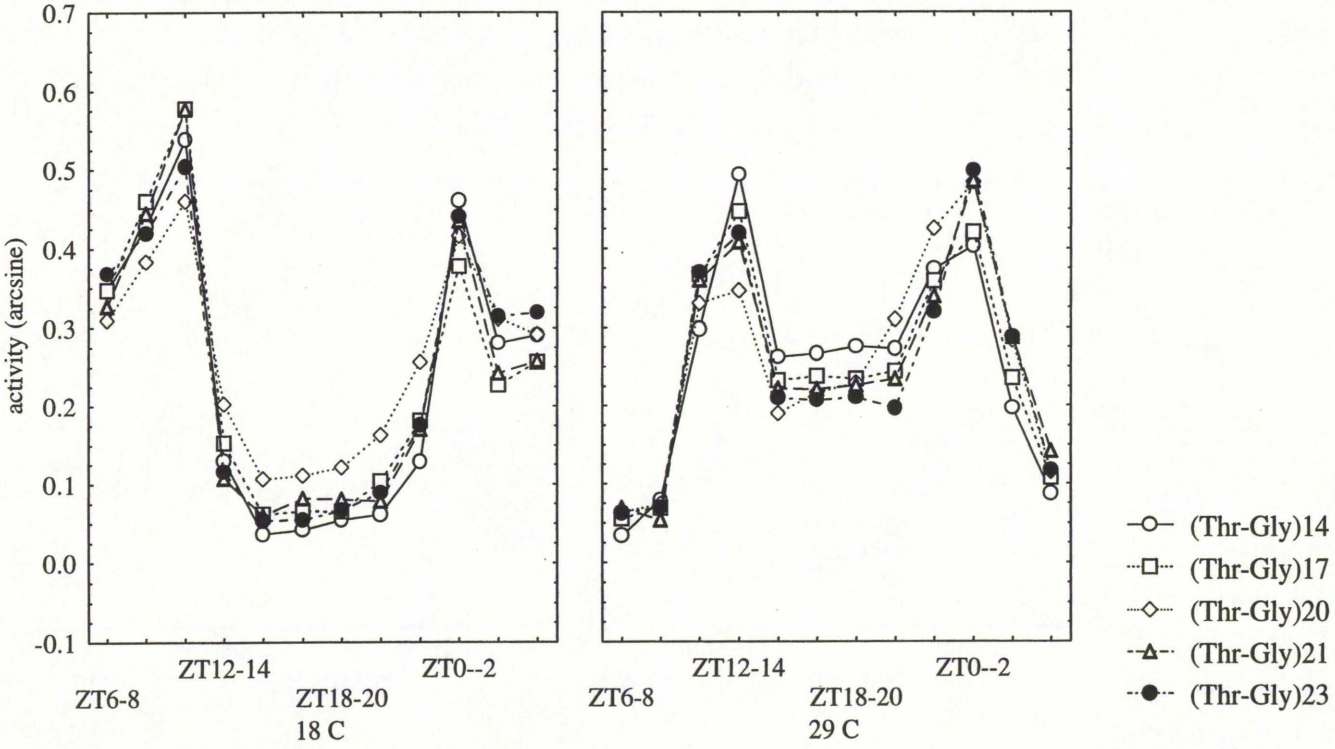
Appendix 6.7 reveals that significantly different relative amounts of activity were observed for the windows spanning ZT20-4 at 29°C and 18°C (except for ZT0-2), indicating a general phenotypic difference between the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles for these particular time bins. When inspecting these two length variants in Fig. 6.7 it is obvious that the shorter length variant is less active at these times than the $(Thr\text{-}Gly)_{20}$. Appendix 6.8 shows the planned comparisons between $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ length variants for each 2h window at different temperatures. Interestingly, no significant differences were observed between the different length alleles at 18°C. In addition the same time windows (ZT20-4) were significantly different at 29°C as found for the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ carriers.

Lecce

A similar analysis was performed for the different Thr-Gly lines of the Lecce population, where Genotype (3 levels), Temperature (2 levels) and Window (12 levels)

Fig. 6.7 Mean transformed proportions of activity of different Thr-Gly length variants, 3-way interaction, Conselve

F(44,2952)=3.47; p<0.0000*



were the main effects investigated. Significant effects were observed for Windows, Window Temperature interaction (W*T) and the Temperature, Genotype, and Window interaction (T*G*W) was marginally significant, again indicating the different Thr-Gly length variants showed dissimilar activity distributions with temperature (Table 6.4).

TABLE 6.4 ANOVA of the Thr-Gly alleles (transformed proportions, *, p<0.05)

LECCE				
EFFECT	MS	df	F	p
GENOTYPE(G)	0.01	3	1.48	0.218
TEMP(T)	0.04	1	3.79	0.052
WINDOW(W)	2.55	11	266.59	0.000*
T*G	0.00	3	0.20	0.898
W*T	0.92	33	96.76	0.000*
T*G*W	0.01	33	1.44	0.050*
ERROR	0.009	2316		

Planned comparisons between the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles revealed a more conservative effect than for the Conselve population, (Appendix 6.9). However, it was interesting to observe differences in activity for lights off (9pm, ZT12) and lights on (9am, ZT0) time periods at both temperatures which included some of the 2h windows which were significantly different for the Conselve lines. The planned comparisons between $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ variants were even more conservative with only two 2h bins at 18°C being significantly different between the genotypes (ZT4-8, Appendix 6.10). Figure 6.8 shows the activity patterns of the Lecce Thr-Gly length variant lines at the two different temperatures. As with the Conselve lines, they all show the same general activity profile, but the $(Thr\text{-}Gly)_{17}$ allele is relatively less active than the $(Thr\text{-}Gly)_{20}$ carrier at most of the 2h windows.

Figures 6.9 and 6.10 take a closer look at the activity profiles of the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ of Conselve and Lecce populations, respectively. Significant differences observed in planned comparisons (Appendix 6.9 and 6.10) are represented with an asterisk.

In summary, it was revealed through these ANOVA's and planned comparisons that the different Thr-Gly length variants subtly behave in a dissimilar way. More specifically it appears that for both Lecce and Conselve a few windows (ZT22-0, 0-2, 2-4, 8-10, 10-12,

Fig. 6.8 Mean transformed proportions of activity of different Thr-Gly length variants, 3-way interaction, Lecce
F(33,2316)=1.44; p<0.0502*

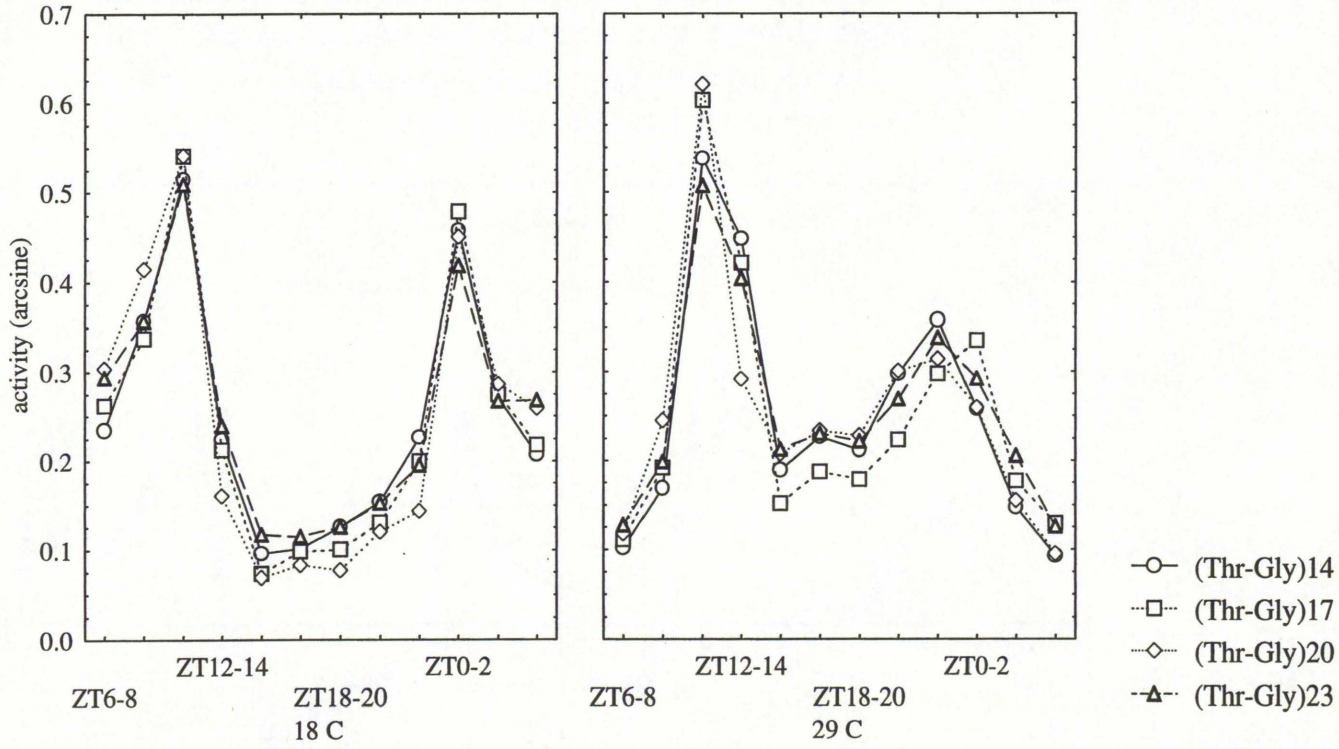


Fig. 6.9 Mean $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ activity profiles
transformed proportions, 3-way interaction, Conselve
$F(11,1452)=4.48$; $p<0.0000$*
(* denotes $p<0.05$)

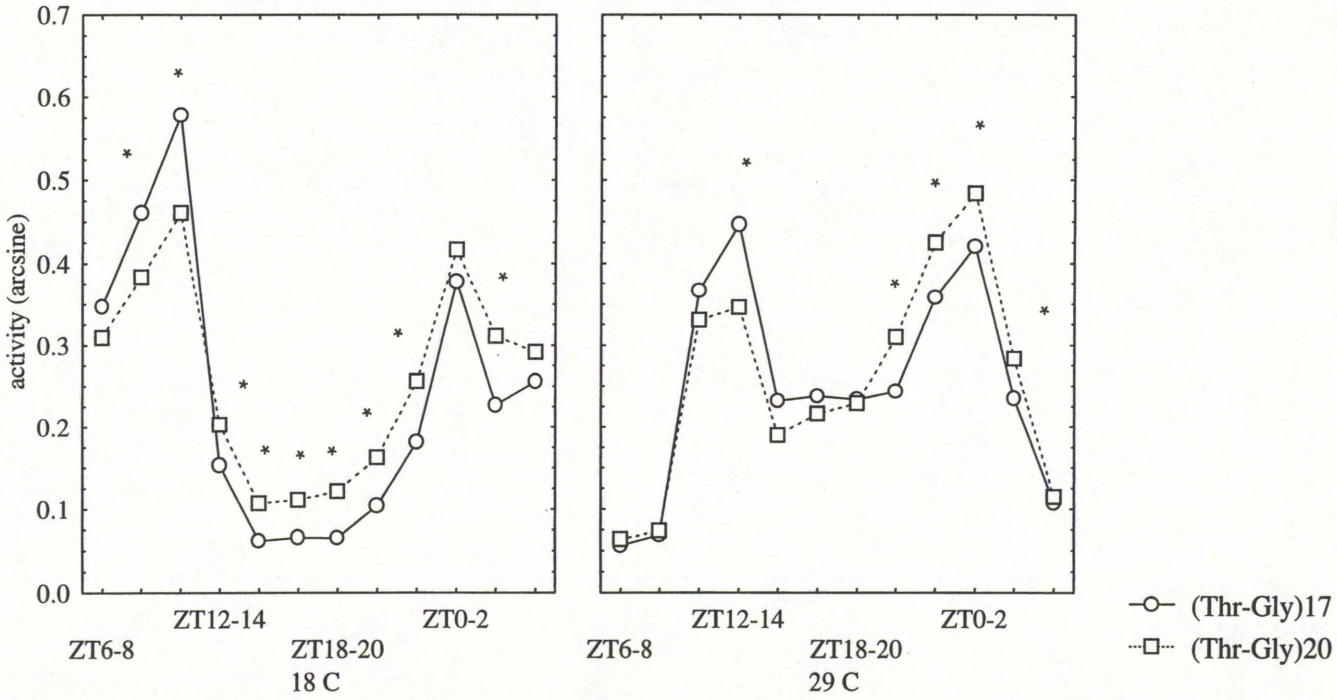
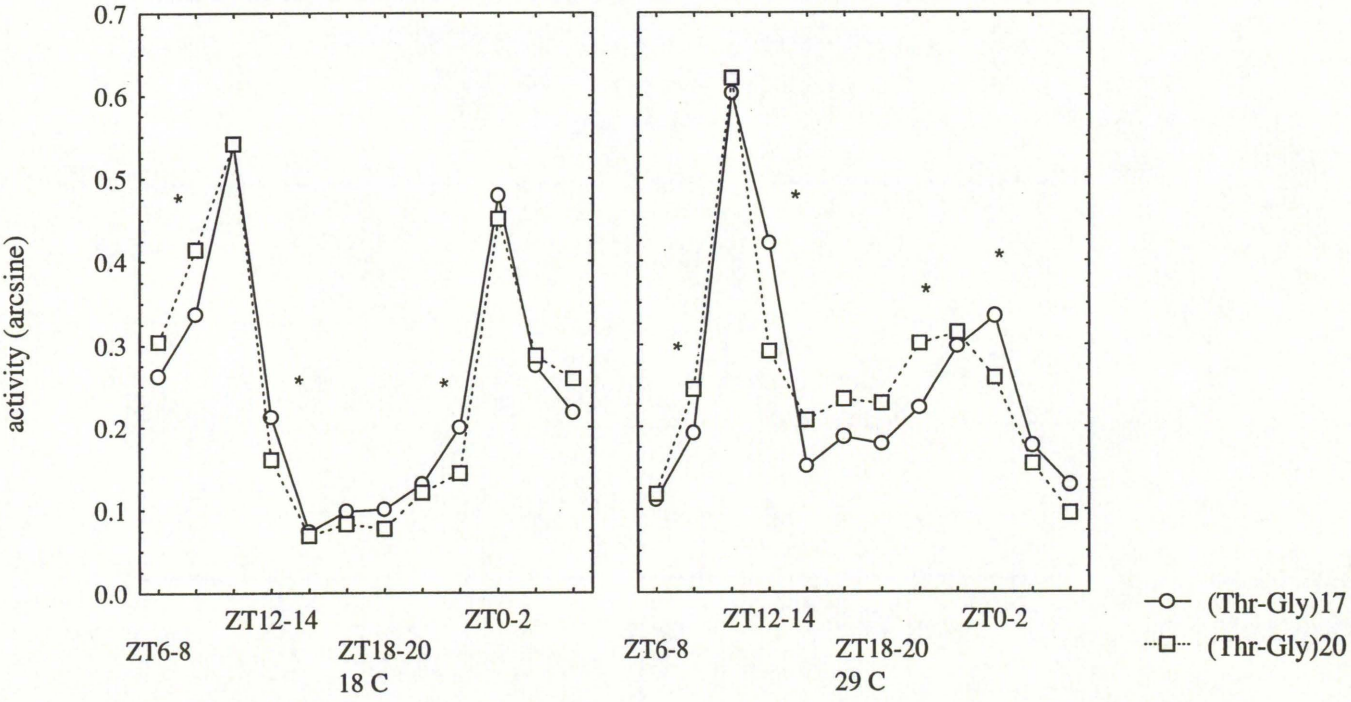


Fig. 6.10 Mean $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ activity profiles

transformed proportions, 3-way interaction, Lecce

$F(11,1308)=2.94$; $p<0.0008$*

(* denotes $p<0.05$)



12-14,) are significantly different for the common length variants, the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$. As the ANOVA was performed on proportions which are interdependent, some caution must be expressed about these conclusions.

6.3.4 Correlations between transformed proportions of each 2h window with Thr-Gly length at the two temperatures

Pearson product moment correlations were also performed to investigate whether there was any linear relationship between Thr-Gly length and relative activity at each 2h window for each population. In this situation, the proportion of activity for each 2h window is independent of that for every other fly in the analysis and so this bypasses any statistical objections.

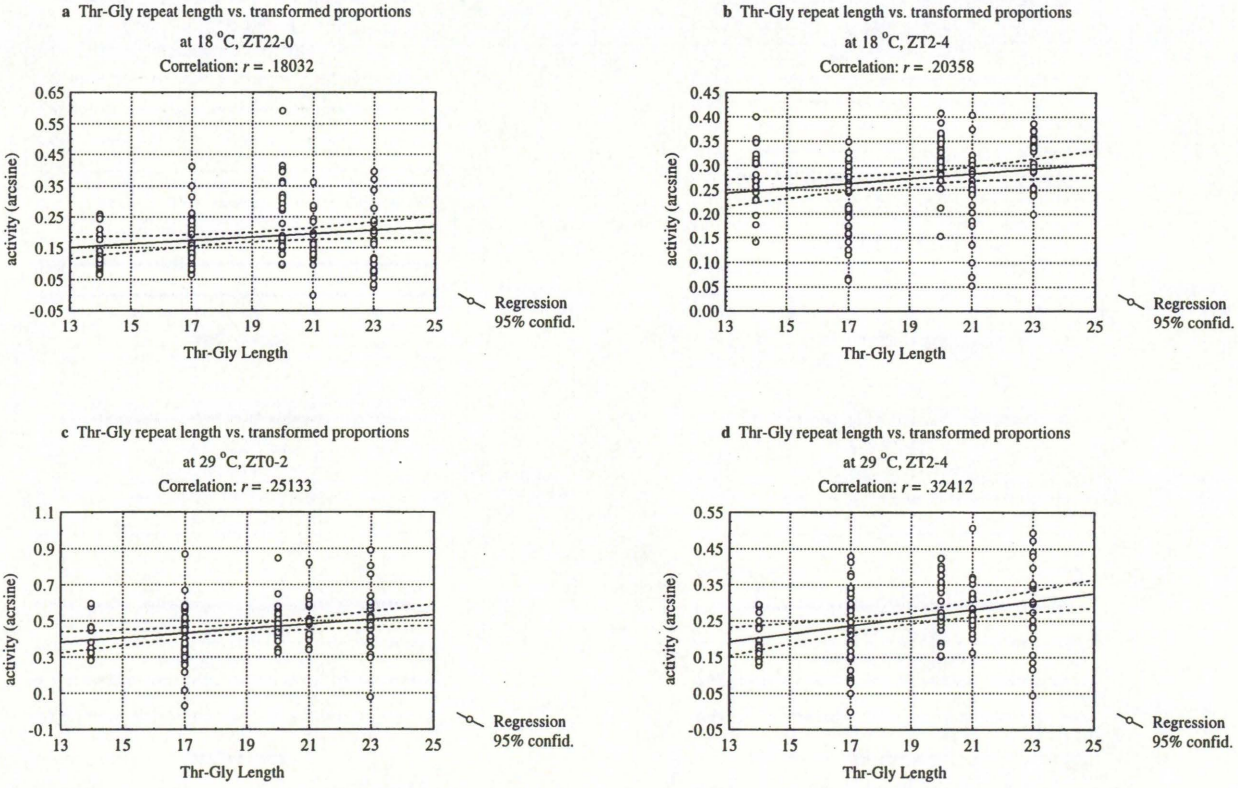
Conselve

The consecutive windows, ZT22-0 and ZT0-2, were significantly correlated with Thr-Gly length at 18°C (Appendix 6.11, Fig. 6.11a and b), and at 29°C significant correlations were observed for ZT12-14, ZT16-20 and ZT0-4 windows (Fig 6.11c and d), the latter two of these is shown in Fig. 6.7. All daytime (ZT0-4) correlations were positive suggesting an increase in Thr-Gly length could be involved in determining the quantity of activity within specific windows, particularly the ZT2-4 window, while the night time (ZT12-14 and ZT16-20) correlations were negatively correlated, suggesting the longer length alleles are relatively less active at night than the shorter length variants.

Lecce

The ZT4-6 and ZT6-8 were significantly correlated with Thr-Gly length at 18°C (Appendix 6.12). In addition, not a single significant correlation was found at 29°C, with the exception of ZT2-4 which was marginal ($p = 0.051$)

Therefore at colder temperatures, there is some evidence that in the afternoons, there is a significant positive correlation between Thr-Gly length and activity. At warmer temperatures, the correlation is restricted to the early daylight hours, ZT0-4. Thus, there is some evidence in both populations that the longer length variants are more relatively active between ZT0-4, at 18°C than the shorter length variants.



Figures 6.11 a to d, CONSELVE

Correlations of the transformed proportions of activity for particular 2h windows at 18 °C and 29 °C

6.4 DISCUSSION

In summary this behavioural study of the Thr-Gly variants of two populations has in LD cycles revealed that they differ in their activity profiles subtly over different time periods. However, this is more true for the Conselve variants than those from Lecce.

More specifically, it is the mid morning (ZT0-4) differences seen between the two most common length variants, the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ which is the most interesting and perhaps the most informative. These differences are supported by the significant positive correlations found between the Conselve Thr-Gly length variants and locomotor activity within the same 2h windows (Figs. 6.11a to d, 6.5 to d and 6.6 and b). Therefore, shorter length variants were less active during this part of the day than the longer ones. This could prove advantageous to flies possessing the $(Thr\text{-}Gly)_{17}$ allele, because they would become less active during the hottest part of the day, perhaps enabling them to avoid desiccation. This could conceivably contribute to an explanation for the southern predominance of the $(Thr\text{-}Gly)_{17}$ allele in the European cline. These characteristic LD phenotypes perhaps should have been stronger with the Lecce variants, as they represent a more southerly Italian population than Conselve. The only evidence for this in the Lecce population is the marginally significant correlation at ZT2-4 (Appendix 6.12). It was also clear that the Conselve $(Thr\text{-}Gly)_{17}$ variant was more active in the hours before darkness (ZT8-12, Fig. 6.9) and less in the hours before dawn until noon (ZT20-4, Fig. 6.9) in comparison to the $(Thr\text{-}Gly)_{20}$ carrier. However, this is barely supported by the Lecce Thr-Gly length variants (Fig. 6.10). Under hot conditions, the greater the activity in the cooler hours before dark and the less in the hotter parts of the day might represent an adaptive strategy for the $(Thr\text{-}Gly)_{17}$ length variant. However, again this is only seen with the Conselve lines. A reciprocal adaptive strategy might also be involved, since there is consistency between the Conselve and Lecce populations, in that at 18°C the $(Thr\text{-}Gly)_{20}$ variant is more active during ZT0-4 than the $(Thr\text{-}Gly)_{17}$ allele suggesting an adaptation to cooler climates (Figs. 6.2, 6.4, 6.9 and 6.10). Perhaps then an adaptive strategy is characteristic of the $(Thr\text{-}Gly)_{20}$ carrier, which is found in predominantly colder climates, for example, northern Europe and southern Australia.

Although these observations are not replicated in both natural populations, they may be real because *D. melanogaster* transformants carrying a single Thr-Gly repeat ((Thr-

$Gly)_1$) show a drastic reduction in the amount of activity compared to $(Thr\text{-}Gly)_{20}$ transformants between ZT2-7 at 18°C, and this is extended at 29°C from ZT2-10 (J.M. Hennesy, pers comm.). An even more bimodal activity pattern is produced for the transformant whose Thr-Gly repeat plus a few flanking amino acids is deleted (Δ(Thr-Gly) , Yu *et al.*, 1987b; J. M. Hennessy pers comm). Consequently these results suggest that the subtle phenotypic differences detected in natural populations are at the edge of a behavioural continuum that can be seen more dramatically with these *in vitro* deleted Thr-Gly variants.

This reduction in locomotor activity of the shorter Thr-Gly length variants at this phase of the day is reminiscent of the 'siesta' which is common in human populations in Mediterranean areas. It is tempting to suggest that in *D. melanogaster*, this apparently analogous response may be an adaptation to the hottest phases of the day, and thereby reflecting a behavioural strategy which attempts to minimise exposure to desiccation. Perhaps then these differences in locomotor behaviour encoded by the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles allied to the difference in temperature compensation discussed in chapter 5, may explain the clinal distribution seen in these European variants. The following hypothesis might be considered, that on hot days, when the $(Thr\text{-}Gly)_{17}$ allele has an optimum 24 h period it might be favoured over the $(Thr\text{-}Gly)_{20}$ variant because of its adaptive 'siesta'. On colder days, the $(Thr\text{-}Gly)_{20}$ allele might be favoured as it is generally more active compared to the $(Thr\text{-}Gly)_{17}$ carrier, furthermore, this also has implications for fitness in terms of the additional opportunities for finding mates or for foraging behaviour under cooler conditions. Thus in regions where there are more 'hot' than 'cold' days such as in southern Europe, the $(Thr\text{-}Gly)_{17}$ variant will be favoured, and vice versa in northern Europe.

The Thr-Gly repeat of *D. melanogaster* may therefore play a role in determining the daily pattern of locomotor activity, when the daily proportion of activity in different phases of the day is taken into account as representing the 'energy' expenditure of the fly. Combined with the observation that this region also appears also to be an important component of the fly's circadian temperature compensation, leads me to propose a selective explanation for the clinal pattern of the Thr-Gly length variation seen in Europe.

CHAPTER 7

7.1 INTRODUCTION

Chapter 4 described the spatial differentiation of the length polymorphism in the Thr-Gly repeat region of the *per* gene in natural Australian populations of *D. melanogaster*. The results revealed that the $(Thr\text{-}Gly)_{20}$ variant showed a significant spatial patterning that was similar to that found in Europe for the same Thr-Gly allele. However the $(Thr\text{-}Gly)_{17}$ in Australia did not give the clinal pattern that was observed in Europe (Costa *et al.*, 1992). The behavioural analysis of European Thr-Gly variants revealed that this polymorphism has an effect on temperature compensation. Generally the shorter length variants were more temperature sensitive with respect to period length then the longer variants. It would therefore be valuable to see whether the Australian Thr-Gly length variants behaved in the same way. In addition the circadian rhythms of rarer length alleles such as the $(Thr\text{-}Gly)_{18}$ (a $(Thr\text{-}Gly)_{20}$ length allele lacking the $(Thr\text{-}Gly)_2$ repeat downstream of the main repeat) and the $(Thr\text{-}Gly)_{22}$, both of which were found in Australia (chapter 3 and 4), had not been analysed before. I have performed a second freerunning behavioural analysis of many Australian Thr-Gly length variants at three different temperatures, 18°C, 25°C and 29°C. These experiments were extended to include 25°C data as current work in the laboratory described in chapter 5 (Fig. 5.16) reveals that the circadian periods of the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ transformants show a distinctive pattern over this temperature range, increasing from 18°C to 25°C and then showing a plateau, or decreases in period length between 25°C and 29°C (Fig. 5.16)

7.1 METHODS

The behavioural analysis presented in this chapter is performed on descendants of flies collected by Linda Partridge and Avis James on the east coast of Australia in February 1993 (see map chapter 4, Fig 4.1). Not all locations were represented; two northern Australian populations, Innisfall Farm (INN) and High Falls Farm (MO) from outside Cairns; two central Australian populations Agnus Farm (AG) near Hervey Bay and Lazy Harry's Farm (LHF) near Yeppoon; and two southern Australian populations collected outside Melbourne, Chappies Farm (M) and Hastings Farm (H) were analysed.

Males identified from the original isofemale lines by PCR which carried the $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{18}$, $(Thr\text{-}Gly)_{20}$, $(Thr\text{-}Gly)_{21}$, $(Thr\text{-}Gly)_{22}$ and $(Thr\text{-}Gly)_{23}$ length repeats (see chapter 2) were crossed to attached-*X ywf* females as described in chapter 5. This provided a self-perpetuating stock of the different length variants. The behavioural experiments were carried out in constant darkness at 18°C, 25°C and 29°C as described in chapter 2.

Statistical analysis of the data was carried out using Minitab version 8 software and Statistica (Statsoft) for windows.

7.3 RESULTS

Tables 7.1 and 7.2 give the mean freerunning periods obtained at 18°C, 25°C and 29°C for flies carrying the $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{18}$, $(Thr\text{-}Gly)_{20}$, $(Thr\text{-}Gly)_{21}$, $(Thr\text{-}Gly)_{22}$ and $(Thr\text{-}Gly)_{23}$ length variants analysed by autocorrelation and spectral procedures respectively. Males from 31 lines from the six different origins were tested, which included one $(Thr\text{-}Gly)_{14}$ allele, eight $(Thr\text{-}Gly)_{17}$ alleles, two $(Thr\text{-}Gly)_{18}$ variants, eleven $(Thr\text{-}Gly)_{20}$ alleles, two $(Thr\text{-}Gly)_{21}$ variants, two $(Thr\text{-}Gly)_{22}$ alleles and five $(Thr\text{-}Gly)_{23}$ variants.

The *a priori* hypothesis was that flies carrying different Thr-Gly length variants would show differences in the temperature sensitivity of their endogenous clock.

7.3.1 Global ANOVA of the Thr-Gly length variants pooled across genotype

A global two way ANOVA was initially performed by pooling all individuals within genotype, giving Genotype (7 levels) and Temperature (3 levels) as the main effects,

AUTOCORRELATION (significance * p<0.05; df, 2, 2227, MS error 0.366)

	line	N/S	genotype	mean period @ 18°C	n @ 18°C	mean period @ 25°C	n @ 25°C	mean period @ 29°C	n @ 29°C	F	p-level	latitude degrees
1	INN	N	14	23.90	29	24.15	31	23.88	21	1.71	0.182	16.88
2	INN	N	17A	24.03	35	23.99	34	23.63	20	3.14	0.043*	16.88
3	INN	N	17B	23.96	27	24.04	14	24.00	23	0.07	0.933	16.88
4	MO	N	17	24.14	28	24.47	19	23.39	19	1.95	0.142	16.88
5	AG	C	17A	23.63	12	24.29	29	24.65	30	12.40	0.000*	25.55
6	AG	C	17B	24.04	14	24.00	25	24.05	33	0.04	0.959	25.55
7	LHF	C	17A	23.23	11	24.15	27	23.84	31	9.10	0.000*	23.13
8	LHF	C	17B	24.27	11	24.31	24	24.21	26	0.18	0.839	23.13
9	H	S	17	23.89	23	24.32	25	23.62	30	9.26	0.000*	38.23
10	M	S	18A	23.57	21	23.71	24	23.99	42	3.79	0.023*	37.68
11	M	S	18B	23.33	23	23.73	26	23.91	28	6.06	0.002*	37.68
12	INN	N	20	24.29	31	24.47	16	23.89	23	4.92	0.007*	16.88
13	MO	N	20	23.87	36	24.13	31	24.24	21	2.94	0.053	16.88
14	AG	C	20A	23.28	9	24.00	23	24.39	28	11.86	0.000*	25.55
15	AG	C	20B	23.83	15	23.77	22	23.61	38	0.99	0.372	25.55
16	LHF	C	20A	23.83	12	24.33	21	24.34	28	3.34	0.036*	23.13
17	LHF	C	20B	23.43	20	24.28	23	24.05	47	11.58	0.000*	23.13
18	M	S	20	24.05	19	24.38	26	23.97	30	3.55	0.029*	37.68
19	H	S	20A	23.76	25	24.39	27	24.27	31	7.95	0.000*	38.23
20	H	S	20B	23.74	23	24.20	23	24.08	30	3.61	0.027*	38.23
21	H	S	20C	23.48	25	24.08	26	23.97	29	7.04	0.001*	38.23
22	H	S	20D	23.56	18	23.88	21	23.78	27	1.45	0.235	38.23
23	INN	N	21	23.52	28	23.95	33	24.40	25	14.06	0.000*	16.88
24	AG	C	21	23.54	13	24.17	27	23.83	32	5.17	0.006*	25.55
25	LHF	C	22A	23.38	13	24.19	24	24.32	33	11.49	0.000*	23.13
26	LHF	C	22B	23.42	13	23.88	26	23.65	20	2.64	0.072	23.13
27	INN	N	23	24.37	30	23.98	26	23.92	25	4.52	0.011*	16.88
28	MO	N	23A	23.87	31	24.23	30	23.91	16	3.09	0.046*	16.88
29	MO	N	23B	24.22	29	24.11	35	23.93	20	1.45	0.234	16.88
30	M	S	23A	23.89	22	24.28	25	23.78	29	4.99	0.007*	37.68
31	M	S	23B	23.71	19	24.22	25	23.90	31	4.06	0.017*	37.68

TABLE 7.1 The mean free running locomotor activity periods of Australian flies carrying different Thr-Gly length variants at 18°C, 25°C and 29°C. Population abbreviations are as follows, Innisfall Farm (INN), High Falls Farm (MO), Cairns; Agnus Farm (AG), Hervey Bay and Lazy Harry's Farm (LHF), Yeppoon; Chappies Farm (M) and Hastings Farm (H), Melbourne. Planned comparison results, latitudes and an arbitrary grouping (see text) of northern (N), southern (S) and central (C) populations for each line are also shown. The Genotype column refers to the Thr-Gly length of the lines. A, B, C and D codes do not refer to different isolength alleles but to different lines within genotype.

SPECTRAL(significance * p<0.05; df 2, 2227, MS error 0.363)

	line	N/S	genotype	mean period @ 18°C	n @ 18°C	mean period @ 25°C	n @ 25°C	mean period @ 29°C	n @ 29°C	F	p -level	latitude degrees
1	INN	N	14	24.21	29	24.26	31	24.24	21	0.05	0.955	16.88
2	INN	N	17A	23.94	35	23.93	34	23.94	20	0.01	0.996	16.88
3	INN	N	17B	23.92	27	23.78	14	23.96	23	0.39	0.685	16.88
4	MO	N	17	24.32	28	24.68	19	24.43	19	2.03	0.131	16.88
5	AG	C	17A	23.92	12	24.38	29	24.78	30	9.32	0.000*	25.55
6	AG	C	17B	24.31	14	23.95	25	24.23	33	2.50	0.082	25.55
7	LHF	C	17A	23.43	11	24.25	27	23.81	31	8.22	0.000*	23.13
8	LHF	C	17B	24.31	11	24.62	24	24.20	26	3.11	0.005*	23.13
9	H	S	17	24.03	23	24.44	25	23.61	30	13.02	0.000*	38.23
10	M	S	18A	23.44	21	23.72	24	23.92	42	4.44	0.012*	37.68
11	M	S	18B	23.37	23	23.84	26	23.93	28	4.76	0.009*	37.68
12	INN	N	20	24.50	31	24.54	16	24.25	23	1.51	0.220	16.88
13	MO	N	20	23.96	36	24.17	31	24.34	21	2.88	0.056	16.88
14	AG	C	20A	23.29	9	24.13	23	24.52	28	14.38	0.000*	25.55
15	AG	C	20B	23.53	15	23.98	22	23.74	38	2.53	0.080	25.55
16	LHF	C	20A	23.40	12	24.40	21	24.67	28	18.79	0.000*	23.13
17	LHF	C	20B	23.59	20	24.51	23	24.06	47	12.55	0.000*	23.13
18	M	S	20	24.14	19	24.49	26	23.98	30	5.24	0.005*	37.68
19	H	S	20A	23.77	25	24.57	27	24.32	31	11.91	0.000*	38.23
20	H	S	20B	24.13	23	24.23	23	23.98	30	1.22	0.296	38.23
21	H	S	20C	23.65	25	24.17	26	24.04	29	5.14	0.006*	38.23
22	H	S	20D	23.63	18	24.10	21	23.72	27	3.62	0.027*	38.23
23	INN	N	21	23.51	28	24.09	33	24.21	25	10.65	0.000*	16.88
24	AG	C	21	23.37	13	24.25	27	23.87	32	9.55	0.000*	25.55
25	LHF	C	22A	23.41	13	24.16	24	24.51	33	15.67	0.000*	23.13
26	LHF	C	22B	23.34	13	23.85	26	23.62	20	3.06	0.047*	23.13
27	INN	N	23	24.43	30	23.99	26	24.16	25	3.80	0.022*	16.88
28	MO	N	23A	24.11	31	24.29	30	23.84	16	2.94	0.053	16.88
29	MO	N	23B	24.31	29	24.14	35	24.15	20	0.72	0.488	16.88
30	M	S	23A	23.70	22	24.13	25	23.72	29	4.12	0.016*	37.68
31	M	S	23B	23.89	19	24.17	25	23.86	31	2.08	0.125	37.68

TABLE 7.2 The mean free running locomotor activity periods of Australian flies carrying different Thr-Gly length variants at 18°C, 25°C and 29°C. Population abbreviations are as follows, Innisfall Farm (INN), High Falls Farm (MO), Cairns; Agnus Farm (AG), Hervey Bay and Lazy Harry's Farm (LHF), Yeppoon; Chappies Farm (M) and Hastings Farm (H), Melbourne. Planned comparison results, latitudes and an arbitrary grouping (see text) of northern (N), southern(S) and central (C) populations for each line are also shown. The Genotype column refers to the Thr-Gly length of the lines. A, B, C and D codes do not refer to different isolength alleles but to different lines within genotype.

followed by planned comparisons of each genotype for autocorrelation and spectral data respectively.

Autocorrelation

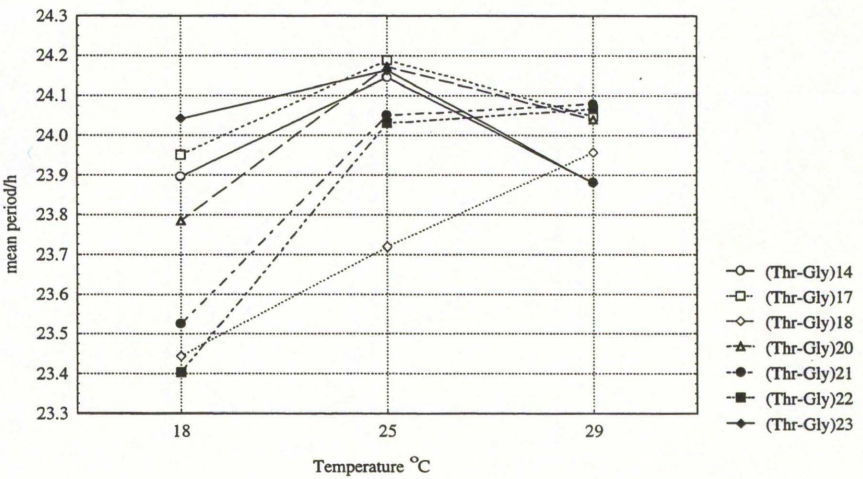
The autocorrelation data gave extremely significant main effects for Genotype, Temperature and their interaction (T*G), see Table 7.2.

Table 7.2 Two Way Fixed ANOVA, Autocorrelation (*, p<0.05)

Source	MS	df	F	p
Genotype(G)	3.44	6	8.57	0.000*
Temperature(T)	12.87	2	32.04	0.000*
T*G	1.81	12	4.51	0.000*
Error	0.402	2299		

Planned comparisons of the pooled Thr-Gly length variants (Appendix 7.1) showed all length alleles, apart from the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ variants, to have significantly different length periods between 18°C and 25°C (Fig 7.1, Appendix 7.1). The $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{18}$, $(Thr\text{-}Gly)_{20}$, and $(Thr\text{-}Gly)_{23}$ length variants showed significant differences between 25°C and 29°C.

Fig 7.1 Plot of Mean Periods of the Australian Thr-Gly Length Variants
2-way interaction (autocorrelation)
F(12,2299)=4.51; p<0.0000*



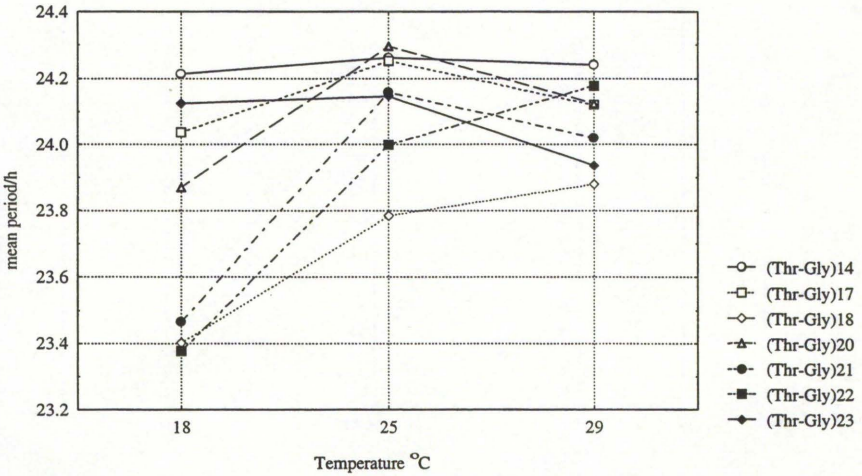
All length variants appeared to follow the general rule of maintaining or decreasing their period length between 25°C and 29°C, apart from the $(Thr\text{-}Gly)_{18}$ (Fig 7.1), which lacks the downstream $(Thr\text{-}Gly)_2$ repeat. The single $(Thr\text{-}Gly)_{14}$ length allele appeared to be the most efficient in temperature compensation, even though the differences between the means (Appendix 7.1) did not differ greatly between this allele and the $(Thr\text{-}Gly)_{17}$ allele In general the most temperature sensitive genotypes between 18°C and 25°C appeared to be the $(Thr\text{-}Gly)_{22}$, $(Thr\text{-}Gly)_{21}$, $(Thr\text{-}Gly)_{20}$ and the $(Thr\text{-}Gly)_{18}$ length alleles. Of these the $(Thr\text{-}Gly)_{20}$, $(Thr\text{-}Gly)_{21}$ and $(Thr\text{-}Gly)_{23}$ carriers stabilise their period between 25°C and 29°C, unlike the $(Thr\text{-}Gly)_{18}$ variant.

Spectral

For Spectral data all main effects were extremely significant when individuals were pooled across genotype (Appendix 7.2). Planned comparisons gave an almost identical result to that with autocorrelation (Appendix 7.3), and the differences between the means (Appendix 7.3, Fig 7.2) showed that the $(Thr\text{-}Gly)_{14}$ allele was the best compensated for period length across temperature, although the flies were representative of one line only. The plateau in period between 25°C and 29°C was observed in all length variants and is also similar to that observed with the different Thr-Gly length transformant flies (J. M. Hennessy personal communication, Fig 5.16). The most temperature sensitive genotypes across 18°C and 29°C were the $(Thr\text{-}Gly)_{18}$, $(Thr\text{-}Gly)_{21}$ and $(Thr\text{-}Gly)_{22}$ alleles. However, the $(Thr\text{-}Gly)_{17}$ length variant was able to maintain its period across these extreme temperatures, whereas its period lengthened at 25°C. The $(Thr\text{-}Gly)_{23}$ length variant showed no significant difference for period length between 18°C and 25°C, however it's period shortened at the higher temperature resulting in a significant difference between 18°C and 29°C and 25°C and 29°C (Fig. 7.2).

This global analysis of period length with respect to temperature shows that different Thr-Gly length variants have different phenotypes at 18°C, 25°C and 29°C, and that significant interactions occur between Thr-Gly length and temperature. Only the single allele of $(Thr\text{-}Gly)_{14}$ appears to have a stable period across temperatures.

Fig. 7.2 Plot of Mean Periods of the Australian Thr-Gly Length Variants
2-way interaction (spectral)
F(12,2299)=5.53; p<0.0000*



7.3.2 ANOVA for all 31 lines

To dissect these results further a two way ANOVA, fixed model, with Lines (31 levels) and Temperature (3 levels) as the main effects was executed, followed again by individual planned comparisons (Table 7.1 and 7.2).

Autocorrelation

Autocorrelation gave significant main and interaction effects (Table 7.3.).

Table 7.3 Two Way Fixed ANOVA, Autocorrelation (*, p<0.05)

Source	MS	df	F	p
Temperature(T)	21.13	2	57.74	0.000*
Line(L)	2.02	30	5.51	0.000*
T*L	1.35	60	3.69	0.000*
Error	0.366	2227		

The planned comparisons, see Table 7.1, showed the single $(Thr\text{-}Gly)_{14}$ length allele to be able to maintain its period across temperature, as did 4 out of 8 $(Thr\text{-}Gly)_{17}$ length alleles (INN17B, MO17, AG17B and LHF17B) , compared with only 3 out of 11 $(Thr\text{-}Gly)_{20}$

length variants (MO20, AG20B and H20D). Both pairs of lines for the $(Thr\text{-}Gly)_{18}$ (M18A and M18B) and $(Thr\text{-}Gly)_{21}$ (INN21 and AG21) length alleles were temperature-sensitive. Only one of the two $(Thr\text{-}Gly)_{22}$ length lines (LHF22B) analysed was significantly different across temperature as were four of the five $(Thr\text{-}Gly)_{23}$ length lines (INN23, MO23A, M23A and M23B).

Spectral

As for autocorrelation all the effects were highly significant (Appendix 7.4). Overall, the spectral data showed that a similar number of lines were significantly different across temperature when compared with autocorrelation data (Table 7.2). The single $(Thr\text{-}Gly)_{14}$ line was efficiently temperature compensated and four out of the eight $(Thr\text{-}Gly)_{17}$ lines tested showed significant temperature effects (AG17A, LHF17A, LHF17B and H17). However seven out of eleven $(Thr\text{-}Gly)_{20}$ lines showed statistical significance (AG20A, LHF20A, LHF20B, M20, H20A, H20C and H20D). Both pairs of lines representing the $(Thr\text{-}Gly)_{21}$ and $(Thr\text{-}Gly)_{22}$ length variants were significantly different over all three temperatures. Of the five lines of $(Thr\text{-}Gly)_{23}$ variants analysed only two showed significant differences, INN23 and M23A.

Figures 7.3a and 7.3b show the mean periods of the individual lines at the three temperatures for autocorrelation and spectral analyses respectively. These are complex figures, therefore the lines will be discussed in the context of their genotype in the following section.

7.3.3 ANOVA on individual genotypes

In order to facilitate the analysis, ANOVAs were performed on each individual Genotype, using Line and Temperature as main effects as in chapter 5 (section 5.3.3). The $(Thr\text{-}Gly)_{14}$ variant was omitted as only one line was available.

Autocorrelation and Spectral

Tables 7.4 and 7.5 show the various ANOVA tables for each genotype. Temperature effects were significant in all comparisons, indicating a general absence of temperature compensation in these lines. The significant Line and interaction effects reveal

Fig. 7.3a Plot of mean locomotor activity periods of all 31 lines
2-way interaction (autocorrelation)
F(60,2227)=3.69; p<.0000*

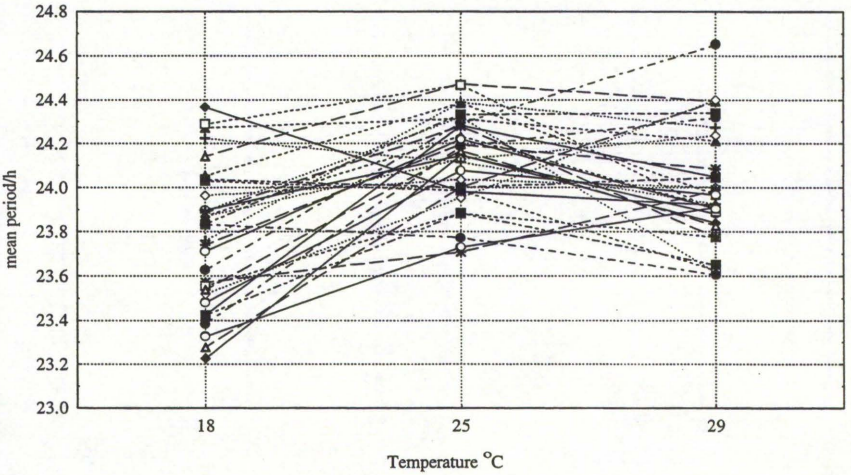
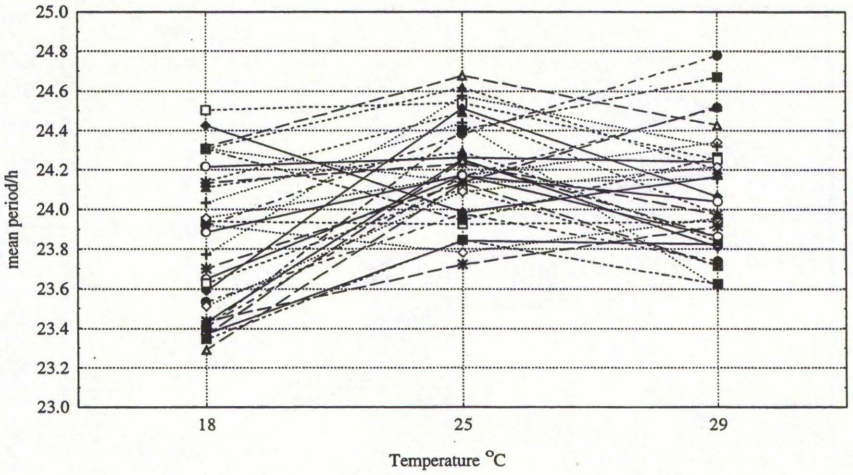


Fig. 7.3b Plot of mean locomotor activity periods of all 31 lines
2-way interaction (spectral)
F(60,2227)=4.44; p<.0000*



Figures 7.3a and b. Means of locomotor activity periods of all 31 lines plotted across temperature for autocorrelation and spectral analyses respectively.

heterogeneity both within and between genotypes. Figures 7.4a to f and 7.5a to f show the mean periods of lines within genotypes at all temperatures for autocorrelation and spectral data respectively. The plots are almost identical between the data types, therefore they will be described together.

The majority of the $(Thr\text{-}Gly)_{17}$ lines (Fig. 7.4a, 7.5a) increase their period length from 18°C to 25°C and then shorten it between 25°C and 29°C. However, a couple of lines do appear to be well temperature compensated between all three temperatures (AG 17B, INN 17B and LHF 17B). Both the $(Thr\text{-}Gly)_{18}$ lines compensate poorly by increasing their period length with an increase in temperature (Fig. 7.4b, 7.5b). The majority of the $(Thr\text{-}Gly)_{20}$ lines lengthen their period between 18°C and 25°C, although they do appear to be better compensated between 25°C and 29°C than the $(Thr\text{-}Gly)_{17}$ lines (Fig 7.4c, 7.5c). In general the $(Thr\text{-}Gly)_{20}$ lines appear to be less heterogeneous than the $(Thr\text{-}Gly)_{17}$ lines. Both the $(Thr\text{-}Gly)_{21}$ lines show poor temperature compensation (Fig 7.4d, 7.5d) as do the $(Thr\text{-}Gly)_{22}$ lines (Fig. 7.4e, 7.5e), although both of these rarer genotypes appear to stabilise their periods better between 25°C and 29°C than between 18°C and 25°C. Three out of the five $(Thr\text{-}Gly)_{23}$ lines increase their period length between 18°C and 25°C and then shorten it between 25°C and 29°C ((MO 23A, (M23 A, and M23 B). The other two lines, INN 23A and MO 23B behave in an opposite direction decreasing their period length between 18°C and 25°C and then increasing between 25°C and 29°C.

These analyses showed that lines within genotypes were heterogeneous in their temperature compensation. The $(Thr\text{-}Gly)_{20}$ lines appeared to be the best compensated between 25°C and 29°C and the $(Thr\text{-}Gly)_{23}$ between 18°C and 29°C. The worst compensation was shown by the $(Thr\text{-}Gly)_{18}$, $(Thr\text{-}Gly)_{21}$ and $(Thr\text{-}Gly)_{22}$ alleles.

A nested ANOVA between the two most common genotypes, the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles was not performed since it required an equal number of lines to be between genotypes which would involve removing at least 25% of the $(Thr\text{-}Gly)_{20}$ variant data.

7.3.4 Pooling across the Thr-Gly lines

The pooled unweighted mean periods across lines were plotted as a function of all three temperatures tested, 18°C, 25°C and 29°C, thus each line contributed one value to

Fig. 7.4a (Thr-Gly)17 lines
2-way interaction (autocorrelation)
F(14,546)=4.16; p<.0000*

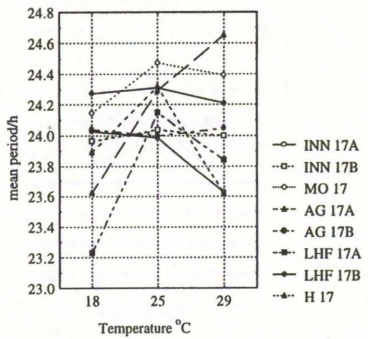


Fig. 7.4b (Thr-Gly)18 lines
2-way interaction (autocorrelation)
F(2,158)=.63; p<.5338

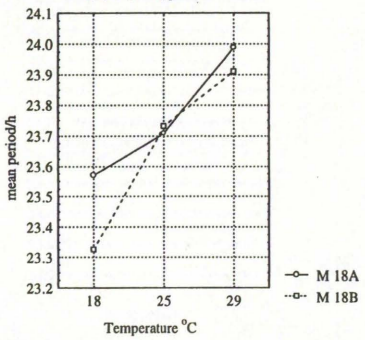


Fig. 7.4c (Thr-Gly)20 lines
2-way interaction (autocorrelation)
F(20,792)=3.45; p<.0000*

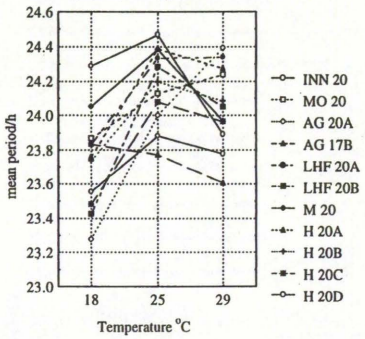


Fig 7.5a (Thr-Gly)17 lines
2-way interaction (spectral)
F(14,546)=4.50; p<.0000*

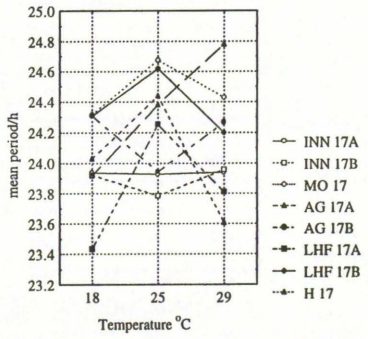


Fig 7.5b (Thr-Gly)18 lines
2-way interaction (spectral)
F(2,158)=.65; p<.5215

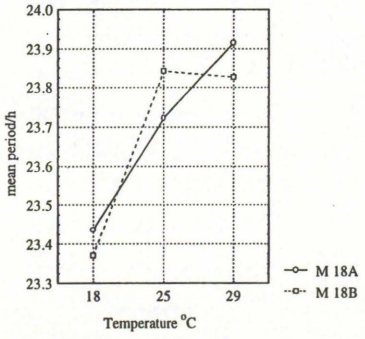
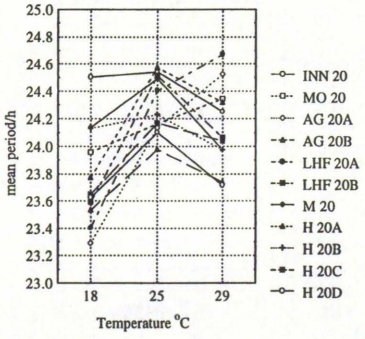
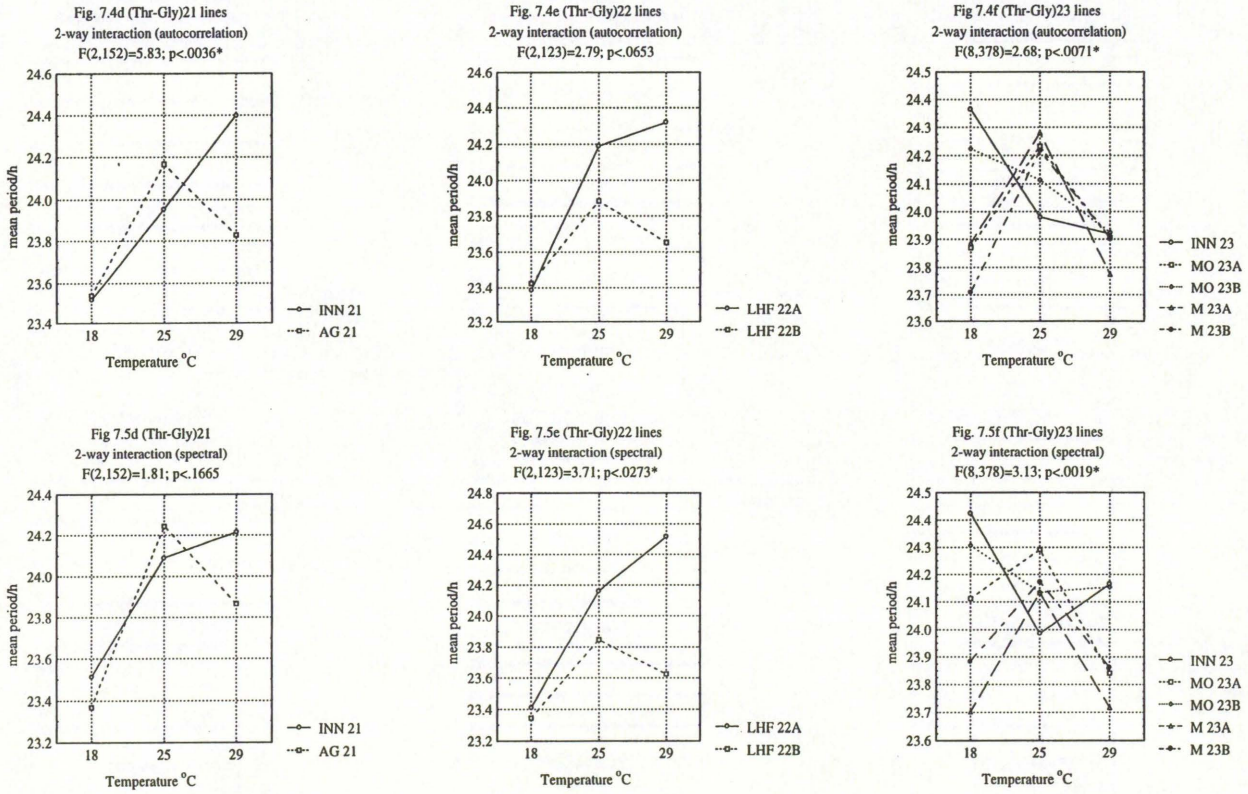


Fig. 7.5c (Thr-Gly)20 lines
2-way interaction (spectral)
F(20,792)=4.66; p<.0000*





Figures 7.4a-f and 7.5a-f. Plots of the mean locomotor activity periods within each genotype for autocorrelation and spectral analyses respectively

the overall mean. Figure 7.6a and 7.6b depict this for autocorrelation and spectral data respectively. In general, both analyses conferred similar results.

Autocorrelation

In this experiment only one $(Thr\text{-}Gly)_{14}$ line was tested which showed excellent temperature compensation between 18°C and 29°C. However, since no further lines were available it is difficult to say whether this result would have been replicated with other $(Thr\text{-}Gly)_{14}$ variants. Moreover some of the European $(Thr\text{-}Gly)_{14}$ lines also showed good temperature compensation (Tables 5.1, 5.2, Figs. 5.7a and 5.8a).

All the variants displayed an approximately 24 h rhythm at the higher temperature of 29°C, and all except the $(Thr\text{-}Gly)_{23}$ allele lengthened their period at this temperature compared to the period lengths found at 18°C (Figure 7.6a). This lengthening of period between 18°C and 25°C was seen with all genotypes.

Taking the mean periods at the extreme temperatures of 18°C and 29°C a difference is observed between the major Thr-Gly length variants compared to the relatively rare length alleles, $(Thr\text{-}Gly)_{18}$, $(Thr\text{-}Gly)_{21}$ and $(Thr\text{-}Gly)_{22}$. The common length alleles ($(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$, and $(Thr\text{-}Gly)_{23}$) which made up almost 91% of the total number of flies sampled, are more efficiently temperature compensated than those which represent only 6% of the natural population. This is observed for the European populations (Fig 5.8a and b) and supports the earlier suggestion that the $(Thr\text{-}Gly)_3$ unit influences temperature compensation. When examining the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length variants it is noticeable that the former has more similar periods at the two extreme temperatures than those found for the latter. However, when the circadian phenotypes of the Australian alleles are compared to the periods found for the equivalent European length alleles with autocorrelation (see Table 7.6 and chapter 5, Figure 5.8a), the $(Thr\text{-}Gly)_{20}$ covers a similar range. The observed periods at 18°C for the $(Thr\text{-}Gly)_{17}$ carrier appear to differ between continents, whereas the periods at 29°C are similar (Table 7.5). The $(Thr\text{-}Gly)_{23}$ length variant however appears to have maintained the same period length range on the two continents, although the small period differences are in opposite directions (Table 7.5).

When comparing the $(Thr\text{-}Gly)_{18}$ and $(Thr\text{-}Gly)_{20}$ length alleles, (differing in the former having no downstream $(Thr\text{-}Gly)_2$ repeat), it is interesting to observe a relatively

Figure 7.6a Mean periods (autocorrelation)

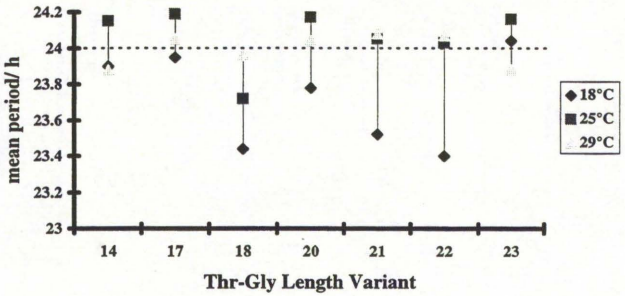


Figure 7.6b Mean periods (spectral)

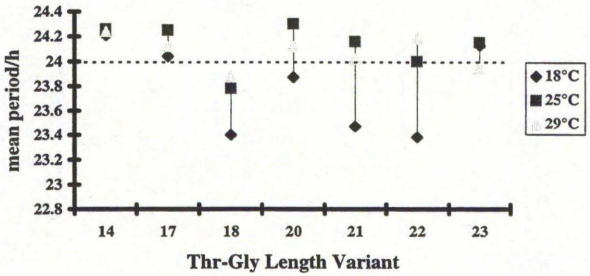


Figure 7.6a and 7.6b Mean locomotor activity periods for Thr-Gly length variants found within Australia (Note the $(Thr\text{-}Gly)_{18}$ variant is a $(Thr\text{-}Gly)_{20}$ lacking the $(Thr\text{-}Gly)_2$ repeat downstream of the main repeat, $(Thr\text{-}Gly)_{20\text{-}2}$).

TABLE 7.4 AUTOCORRELATION

REPEAT LENGTH ALLELES

	$(Thr\text{-}Gly)_{17}$				$(Thr\text{-}Gly)_{18}$				$(Thr\text{-}Gly)_{20}$				$(Thr\text{-}Gly)_{21}$				$(Thr\text{-}Gly)_{22}$				$(Thr\text{-}Gly)_{23.}$			
EFFECT	MS	df	F	p	MS	df	F	p	MS	df	F	p	MS	df	F	p	MS	df	F	p	MS	df	F	p
TEMP(T)	3.46	2	9.36	0.000*	3.35	2	9.92	0.000*	11.1	2	34.1	0.000*	4.28	2	10.7	0.000*	3.82	2	9.50	0.000*	2.48	2	6.32	0.002*
LINE(L)	2.53	7	6.86	0.000*	0.39	1	1.16	0.283	2.00	10	6.13	0.000*	0.46	1	1.14	0.287	2.77	1	6.87	0.010*	0.32	4	0.82	0.511
T*L	1.53	14	4.16	0.000*	0.21	2	0.63	0.533	1.13	20	3.45	0.000*	2.34	2	5.83	0.004*	1.12	2	2.79	0.065	1.05	8	2.68	0.007*
ERROR	0.37				0.34				0.33				0.40				0.40				0.392			

TABLE 7.5 SPECTRAL

REPEAT LENGTH ALLELES

	$(Thr\text{-}Gly)_{17}$				$(Thr\text{-}Gly)_{18}$				$(Thr\text{-}Gly)_{20}$				$(Thr\text{-}Gly)_{21}$				$(Thr\text{-}Gly)_{22}$				$(Thr\text{-}Gly)_{23.}$			
EFFECT	MS	df	F	p	MS	df	F	p	MS	df	F	p	MS	df	F	p	MS	df	F	p	MS	df	F	p
TEMP(T)	2.14	2	5.46	0.004*	3.05	2	11.4	0.000*	15.6	2	44.6	0.000*	6.27	2	12.3	0.000*	4.54	2	9.5	0.000*	1.24	2	4.09	0.018*
LINE(L)	3.77	7	9.64	0.000*	0.01	1	0.02	0.889	2.5	10	7.15	0.000*	0.46	1	0.89	0.346	5.17	1	10.8	0.001*	1.7	4	5.62	0.000*
T*L	1.76	14	4.5	0.000*	0.17	2	0.65	0.521	1.63	20	4.66	0.000*	0.93	2	1.81	0.166	1.77	2	3.71	0.027*	0.95	8	3.13	0.002*
ERROR	0.39				0.27				0.35								0.48				0.302			

Table 7.4 and 7.5 ANOVA on individual genotypes for autocorrelation and spectral periods respectively.
(*, $p < 0.05$).

large difference in the period length between these two genotypes at 18°C, approximately 0.4h. This result is analogous to differences observed between the European $(Thr\text{-}Gly)_{17}$ variant and its equivalent length allele lacking the $(Thr\text{-}Gly)_2$ downstream repeat , the $(Thr\text{-}Gly)_{15}$, which differs in the same direction, (chapter 5, Figure 5.8b). This indicates that the length of the entire Thr-Gly region may be important for temperature compensation.

Table 7.5 Mean locomotor activity periods across lines for the common length alleles for Australia and Europe, Autocorrelation

Thr-Gly Length Variant	Australian mean period/h @ 18°C	Australian mean period/h @ 29°C	European mean period/h @ 18°C	European mean period/h @ 29°C
17	23.95	24.05	23.60	24.06
20	23.78	24.04	23.70	23.90
23	24.04	23.90	23.90	24.02

Generally, in all the variants the periods generally lengthen from 18°C to 25°C then plateau to 29°C (Fig 7.6a), suggesting a threshold-like effect, where past a certain point, changes in the underlying temperature compensation mechanisms are introduced.

Spectral

In agreement with the previously described autocorrelation data the majority of the Thr-Gly length variants show an approximate 24h circadian phenotype at 29°C and all shorten their period with a decrease in temperature, again with the exception of the $(Thr\text{-}Gly)_{23}$ length variants (Fig. 7.6b). The less frequent Australian alleles the $(Thr\text{-}Gly)_{18}$, $(Thr\text{-}Gly)_{21}$ and $(Thr\text{-}Gly)_{22}$ show poorer temperature compensation than the 14-17-20-23 allelic series.

Between 18°C and 29°C (Figure 7.6b), the $(Thr\text{-}Gly)_{17}$ variant genotype efficiently maintains its period as does the $(Thr\text{-}Gly)_{23}$ allele. The $(Thr\text{-}Gly)_{20}$ variant shows slightly more variation between these temperatures. Table 7.6 shows the Australian and European period lengths for the common length variants. The Australian $(Thr\text{-}Gly)_{17}$ allele shows a period closer to 24h than its European cousin and is relatively better temperature compensated. The Australian $(Thr\text{-}Gly)_{20}$ alleles are less well temperature compensated

than the European isolength alleles. The $(Thr\text{-}Gly)_{23}$ length variants however show no obvious differences between the two continental samples (Table 7.6).

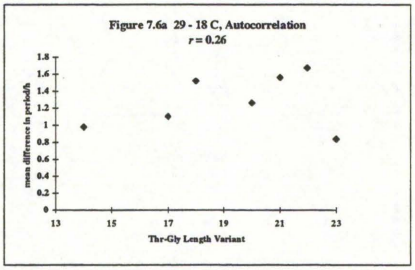
Table 7.6 Mean locomotor activity periods for the common length alleles for Australia and Europe, Spectral

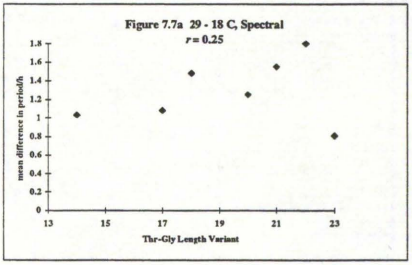
Thr-Gly Length Variant	Australian mean period/h @ 18°C	Australian mean period/h @ 29°C	European mean period/h @ 18°C	European mean period/h @ 29°C
17	24.04	24.10	23.70	23.90
20	23.90	24.10	23.70	23.70
23	24.10	23.90	24.10	23.80

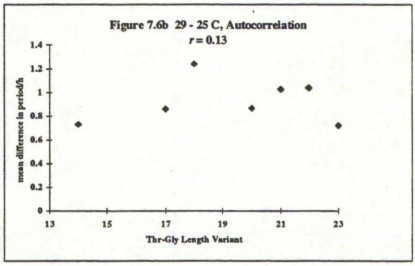
The majority of the Australian Thr-Gly length alleles shorten their circadian period at 18°C , exceptions being the single $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ variant. The only genotype that continues to lengthen its period with temperature is the $(Thr\text{-}Gly)_{22}$ allele, whereas all the other alleles increase their period length at 25°C and then have a shorter period when at 29°C. The $(Thr\text{-}Gly)_{14}$ variant has an excellently maintained period of 24.2h at all three temperatures.

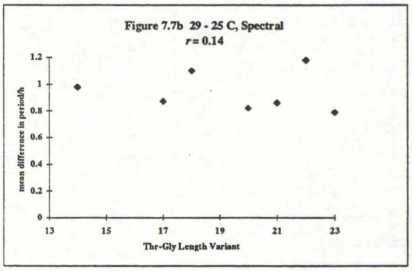
7.3.5 Pooling mean differences of period across lines

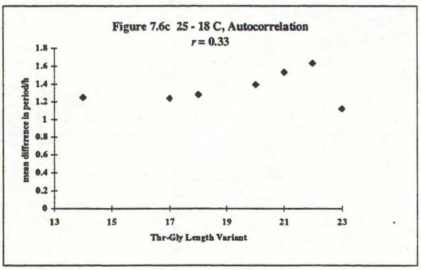
Plotting the difference in mean period between lines for each genotype revealed some interesting observations (Fig. 7.7a, b, c and 7.8a, b, c. These differences were calculated for each line by taking the absolute difference between the two periods at different temperatures (for example, mean period @ 29°C-mean period @ 18°C), always subtracting the period at the lower temperature from the one observed at the higher temperature. These were then averaged within each of the seven individual genotypes, with the addition of +1 to change any negative values to positives ones, thereby providing a correlation.

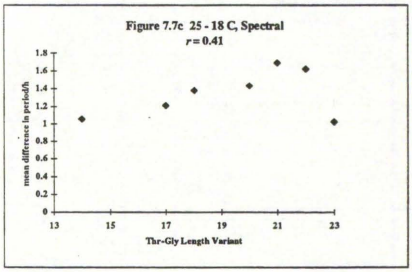












Figures 7.7a, b, c and Figures 7.8a, b, c, Plots of differences in mean locomotor activity periods for Australian Thr-Gly length variants between 18°C and 29°C, 25°C and 29°C and 18°C and 25°C for Autocorrelation and Spectral data respectively.

Autocorrelation and Spectral

As a general observation it was noted that as the Thr-Gly repeat length increased, the difference in mean period did so in a concordant manner, apart from the longest variant the $(Thr\text{-}Gly)_{23}$ (Figures 7.7a, b, c and 7.8a, b and c). This anomaly resulted in no significant Pearson product moment (r) correlations or significant Spearman rank order (r_s) correlations (Appendix 7.5 and 7.6). Observing the common length alleles $(Thr\text{-}Gly)_{14}$, $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ alone, shows an almost homogenous range of mean period differences. These observations are unlike those seen with the European and north African populations.

7.3.6 Individual line differences of locomotor activity rhythms

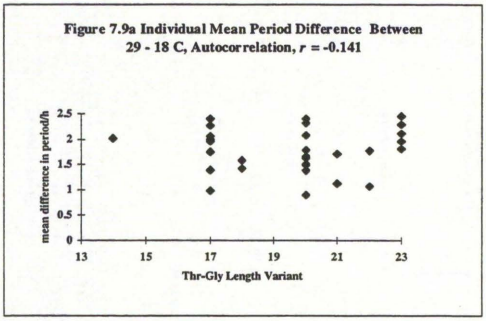
Individual line differences for period between 18°C and 29°C , 25°C and 29°C and 18°C and 25°C were calculated as before in section 7.3.5 and +2 was added to change negative values into positive ones.

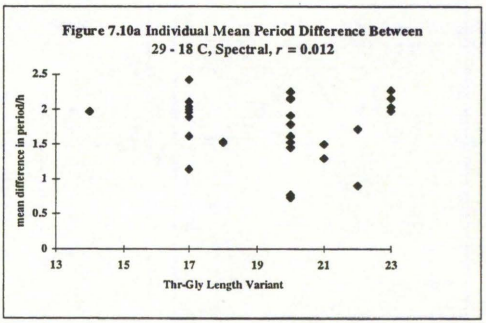
Autocorrelation and Spectral

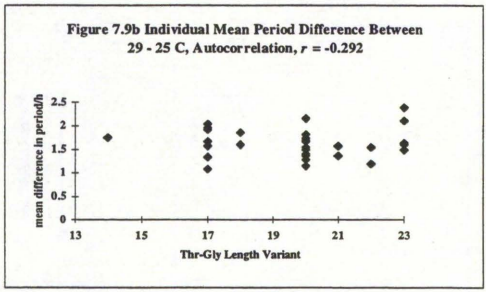
For both sets of data no significant correlation coefficient for either Pearson's product moment (r) or Spearman's rank order (r_s) was observed, see Figures 7.9a, b, c and Figures 7.10a, b, and c for autocorrelation and spectral analysis respectively. The majority of the length alleles have an even distribution of mean differences for period across all temperature comparisons. The noticeable effects are that in the spectral analysis, the 18°C-29°C and 25°C-29°C comparisons show the $(Thr\text{-}Gly)_{20}$ variant to have smaller differences than the $(Thr\text{-}Gly)_{17}$, even though there is considerable overlap in these patterns. This was also observed for the European Thr-Gly length alleles (Figs. 5.10a and b) for period differences between 18°C and 29°C.

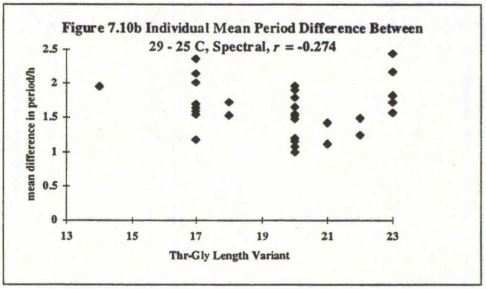
7.3.7 Northern and southern population analysis

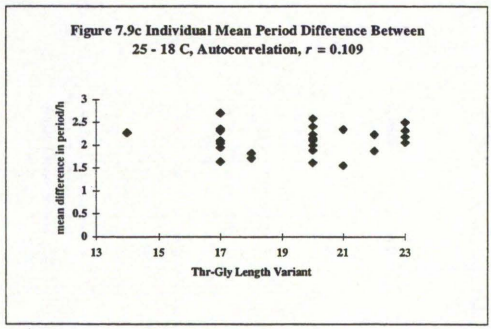
When examining the northern and southern populations of the European and North African flies a significant differences for periods between genotypes of different origin were observed (chapter 5, section 5.3.11). Therefore it was decided to perform a similar analysis on the Australian flies. Flies from populations collected at Innisfall Farm (INN, latitude











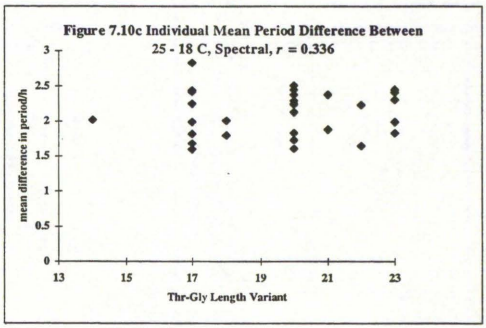


Figure 7.9a, b, c and 7.10a, b,and c. Individual line differences in mean locomotor activity period for each Australian Thr-Gly length variant for autocorrelation and spectral data respectively.

16.88°) and High Falls Farm (MO, latitude 16.88°) were deemed to be northern populations and flies derived from the isofemale line collected at Chappies Farm (M, latitude 37.68°) and Hastings Farm (H, latitude 38.23°) represented southern populations. The centrally located populations Lazy Harry's Farm (LHF) and Agnus Farm (AG) were excluded from this analysis as it was thought that comparing populations from extreme localities might maximise the probability of any significant geographical effects. A three way ANOVA was carried out for all alleles carrying $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length repeats, since the other length alleles represented either northern southern or central groupings exclusively.

Autocorrelation

For autocorrelation data there was a significant Temperature (T), North/South (N/S) and Temperature, North/South interaction effect (T*N/S). This suggests that the northern and southern alleles behave differently at different temperatures, see Table 7.10.1. Also a nearly significant Temperature and Genotype interaction effect (T*G) was observed which tentatively suggests that the 3 most common length alleles are acting differently at different temperatures.

Table 7.7 3-WAY ANOVA for northern and southern populations of $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length alleles (*, $p<0.05$)

AUTOCORRELATION				
EFFECT	MS	df	F	p
GENOTYPE(G)	0.21	2	0.55	0.580
TEMP(T)	8.56	2	22.60	0.000*
N/S	3.36	1	8.86	0.003*
T*G	0.89	4	2.34	0.053
G*N/S	0.06	2	0.16	0.850
T*N/S	2.88	2	7.60	0.001*
T*G*N/S	0.73	4	1.90	0.103
ERROR	0.38	1211		

Planned comparisons were performed (Table 7.8), and this enabled us to observe that, as with the European data there was a significant difference in phenotype between the northern and southern Thr-Gly length alleles. Each of the southern common length alleles was significantly different across all three temperatures whereas their northern neighbours were

not (Table 7.8, Fig. 7.11a and b). Inspection of Fig. 7.11 b reveals the interaction effects of the southern alleles, with the major variants increasing their period from 18°C to 25°C , then decreasing it from 25°C to 29°C . The $(Thr\text{-}Gly)_{17}$ allele shows the largest temperature sensitive phenotype between 25°C and 29°C.

Table 7.8 Planned comparisons between different temperatures for $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length alleles

AUTOCORRELATION (*, $p<0.05$, MS error 0.38).

TEMP /°C	ORIGIN N/S	Thr-Gly LENGTH	MEAN PERIOD/h	planned comparison	MS	df	F	p
18	N	17	24.04	18-25-29	0.31	2	0.81	0.450
25			24.13					
29			24.00					
18	S	17	23.89	18-25-29	3.39	2	8.92	0.000*
25			24.32					
29			23.62					
18	N	20	24.06	18-25-29	0.57	2	1.50	0.220
25			24.24					
29			24.06					
18	S	20	23.71	18-25-29	7.10	2	18.69	0.000*
25			24.20					
29			24.02					
18	N	23	24.15	18-25-29	1.08	2	2.84	0.059
25			24.12					
29			23.92					
18	S	23	23.80	18-25-29	3.01	2	7.93	0.000*
25			24.25					
29			23.84					

Spectral

With spectral circadian periods significant Genotype (G), Temperature (T), Northern/Southern (N/S). Temperature, Genotype interaction(T*G), Temperature, Northern/Southern interaction (T*N/S) and Temperature, Genotype, Northern/ Southern interaction (T*G*N/S) effects were observed, see Table 7.9. The latter observation is the most interesting and again suggests a differentiation between the origins of the fly populations.

Planned comparisons were then performed and as before all southern alleles gave significant temperature effects whereas the northern alleles did not (Table 7.10). This result indicates that the fly populations from northern Australia are more efficient at temperature compensation than the southern populations. These results are illustrated in Figs. 7.12a and b. This graph shows again that the $(Thr\text{-}Gly)_{17}$ alleles from the southern populations have the largest change in period length between 25°C and 29°C.

Figure 7.11a and b, Autocorrelation, Plot of Means

3-way interaction

F(4,1211)=1.93; p<.1035

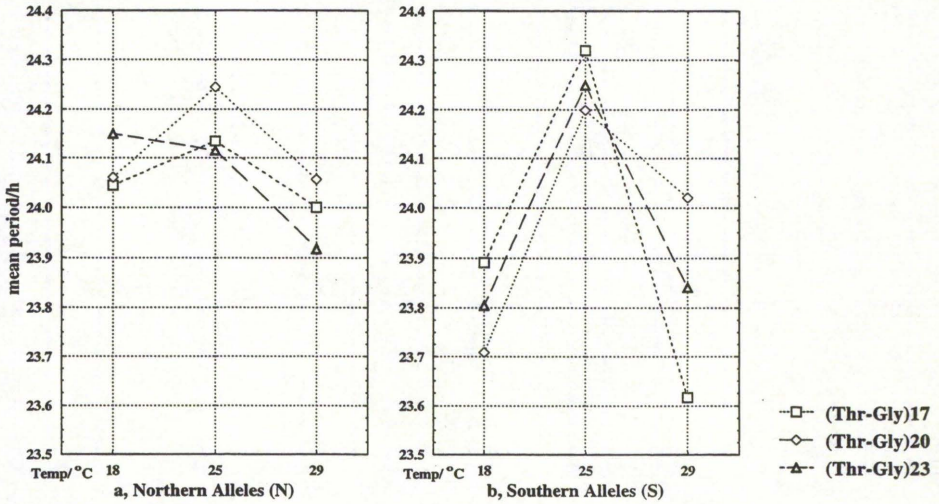
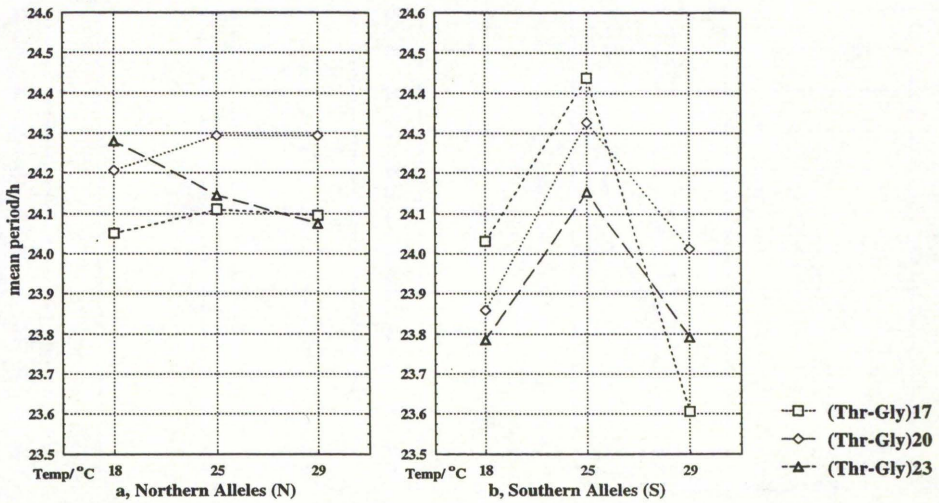


Figure 7.11a and b. Plots of the mean period length of northern and southern length alleles with respect to temperature for autocorrelation data.

Figure 7.12a and b Spectral, Plot of Means

3-way interaction

F(4,1211)=2.50; p<.0410*



Figures 7.12a and b Plots of the mean period length of northern and southern length alleles with respect to temperature for spectral data.

Table 7.9 3-WAY ANOVA for northern and southern populations of $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length alleles (*, $p<0.05$)

SPECTRAL				
EFFECT	MS	df	F	p
GENOTYPE(G)	1.83	2	5.47	0.004*
TEMP(T)	6.12	2	18.31	0.000*
N/S	6.98	1	20.89	0.000*
T*G	0.96	4	2.86	0.022*
G*N/S	0.68	2	2.02	0.130
T*N/S	5.20	2	15.56	0.000*
T*G*N/S	0.84	4	2.50	0.041*
ERROR	0.334	1211		

Table 7.10 Planned comparisons between different temperatures for $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ length alleles

SPECTRAL(*, $p<0.05$, MS error 0.334).

TEMP /°C	ORIGIN N/S	Thr-Gly LENGTH	MEAN PERIOD/h	planned comparison	MS	df	F	p
18	N	17	24.05	18-25-29	0.074	2	0.22	0.798
25			24.11					
29			24.10					
18	S	17	24.03	18-25-29	4.73	2	14.56	0.000*
25			24.44					
29			23.61					
18	N	20	24.21	18-25-29	0.15	2	0.46	0.630
25			24.29					
29			24.29					
18	S	20	23.86	18-25-29	6.73	2	20.13	0.000*
25			24.33					
29			24.01					
18	N	23	24.28	18-25-29	0.83	2	2.50	0.084
25			24.14					
29			24.08					
18	S	23	23.79	18-25-29	2.19	2	6.57	0.002*
25			24.15					
29			23.79					

7.3.8 Northern and southern analysis of $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles

To investigate these differences further a similar ANOVA was performed using the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles only.

Autocorrelation

Significant effects were observed for Temperature (T), Northern/Southern (N/S), Temperature, Genotype interaction (T*G), Temperature, Northern/Southern interaction (T*N/S) and Temperature, Genotype, Northern/Southern interaction, (T*G*N/S) (Table 7.11). The significant T*G and T*G*N/S are particularly important and the differences

between genotype with respect to temperature are reflected in Fig 7.11b, where the (Thr-Gly)$_{17}$ allele shows the biggest difference in period length between 25°C and 29°C.

Table 7.11 3-WAY ANOVA for northern and southern Australian populations of (Thr-Gly)$_{17}$ and (Thr-Gly)$_{20}$ length alleles (*, p<0.05)

AUTOCORRELATION				
EFFECT	MS	df	F	p
GENOTYPE(G)	0.33	1	0.90	0.343
TEMP(T)	5.81	2	15.61	0.000*
N/S	2.55	1	6.85	0.009*
T*G	1.35	2	3.62	0.027*
G*N/S	0.03	1	0.08	0.782
T*N/S	1.50	2	4.01	0.018*
T*G*N/S	1.32	2	3.55	0.029*
ERROR	0.371	824		

Spectral

When performing a 3 way analysis of variance of Spectral data for (Thr-Gly)$_{17}$ and (Thr-Gly)$_{20}$ northern and southern alleles, all effects were significant apart from a Temperature, Northern/Southern interaction effect, see Table 7.12. The significant Genotype, G*T interaction and G*T*N/S interaction agrees with the autocorrelation ANOVA and Fig. 7.12b in highlighting the poorer temperature compensation of the (Thr-Gly)$_{17}$ allele between 25°C and 29°C. This underlines the previous analyses in that the origin of alleles has an important influence on phenotype.

Table 7.12 3-WAY ANOVA for northern and southern Australian populations of (Thr-Gly)$_{17}$ and (Thr-Gly)$_{20}$ length alleles (*, p<0.05)

SPECTRAL				
EFFECT	MS	df	F	p
GENOTYPE(G)	1.80	1	5.19	0.023*
TEMP(T)	4.86	2	14.02	0.000*
N/S	2.49	1	7.19	0.008*
T*G	1.43	2	4.11	0.017*
G*N/S	0.71	1	2.05	0.152
T*N/S	4.00	2	11.54	0.000*
T*G*N/S	1.13	2	3.26	0.039*
ERROR	0.345	824		

7.3.9 Correlations between period length and latitude of Thr-Gly variants at different temperatures

As before in chapter 5 (section 5.3.14), correlations with latitude were performed on individual fly periods for each of the common length variants, $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ at each temperature, 18°C , 25°C and 29°C. Only these length variants were chosen since all others were found in no more than two localities.

Autocorrelation and Spectral

Pearsons *r* and Spearmans r_s were calculated and significant coefficients were observed for the $(Thr\text{-}Gly)_{20}$ allele at 18°C (autocorrelation) . A significant Pearsons *r* for the $(Thr\text{-}Gly)_{17}$ variant at 29°C was also observed (Appendix 7.7, autocorrelation). With spectral data significant correlation coefficients for both Pearsons *r* and Spearmans r_s for the $(Thr\text{-}Gly)_{17}$ at 18°C and 25°C and the $(Thr\text{-}Gly)_{20}$ at 29°C were observed, and also a significant Pearsons *r* for the $(Thr\text{-}Gly)_{17}$ at 29°C (Appendix 7.8). The correlations using the spectral data were not in agreement with those observed with the European alleles. That is to say, firstly the Australian $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles both decreased their period with an increase in latitude similar to the European $(Thr\text{-}Gly)_{20}$. However the European $(Thr\text{-}Gly)_{17}$ variant decreased its period with an increase in latitude.

It is interesting to note that the correlation for the $(Thr\text{-}Gly)_{17}$ allele at 25°C is positive and opposite to that of all the other significant correlation coefficients found.

7.4 DISCUSSION

In summary, this comprehensive investigation of the circadian locomotor phenotype of Australian Thr-Gly length alleles has shown the $(Thr\text{-}Gly)_{20}$ to be slightly less efficient at maintaining its period length between different temperatures than the $(Thr\text{-}Gly)_{17}$. Furthermore, when examining individual lines, a greater number of $(Thr\text{-}Gly)_{20}$ length alleles failed to maintain their circadian rhythm length compared to $(Thr\text{-}Gly)_{17}$ lines when tested across all three different temperatures. When pooling across lines within a genotype there is no correlation between Thr-Gly length and temperature compensation (Figs 7.7a, b, c and Figs. 7.8a, b, c). These results would seem to contradict both those found in chapter 5, where the $(Thr\text{-}Gly)_{17}$ variant showed slightly poorer temperature compensation than the $(Thr\text{-}Gly)_{20}$, and the results of the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ transformant flies (J. M.

Hennessy, personal communication). However, the heterogeneity among lines, as shown by the T*G*N/S interactions (Table 11 and 12), shows a significant difference in temperature compensation between the southern $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles between 25°C and 29°C (Fig 7.11b and 12b).

As northern Australia is situated in the tropic of Capricorn, the monthly temperatures are not changing as dramatically through the year (7-8°C) as in southern Australia (11- 12°C, Gentilli, 1971). However these northern temperatures and of course the humidity are higher than those found in southern Europe. Therefore the northern populations of *D. melanogaster* appear to have evolved a particularly robust temperature compensation mechanism (Fig. 11a and 12a). It is also found that the northern populations of *D. melanogaster* have a period closer to 24h (Fig. 11a and 12a) which would be important at these higher temperatures to assure activity in the cooler times of day (dawn and dusk peaks) and to avoid desiccation and male sterility. Generally the $(Thr\text{-}Gly)_{17}$ allele is found at greater frequencies than the $(Thr\text{-}Gly)_{20}$ allele in the north of Australia (chapter 4) and the former northern allele is better compensated and has a period closer to 24h than the latter (Fig. 7.11a and 12a). This could help explain the northern population structure of these alleles. The south of Australia is not really comparable to northern Europe in terms of latitude and therefore climate, but is more similar to southern Spain and Italy, especially the south east tip of Australia (including Melbourne) which is known to have a Mediterranean climate (reported in Cowling *et al.*, 1996). Therefore the difference in the southern $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles for temperature compensation between 25°C and 29°C (probably the most relevant temperatures due to their generally elevated range in Australia), where the former is less temperature compensated than the latter is interesting. The observation that the frequency of the $(Thr\text{-}Gly)_{20}$ variant is high in southern Australia tallies with the observations in this chapter that the same allele of southern populations is better compensated and has a period closer to 24h. When taken together these observations could explain the higher frequencies of the $(Thr\text{-}Gly)_{20}$ allele in the south.

The three rarer length alleles tested, the $(Thr\text{-}Gly)_{18}$, $(Thr\text{-}Gly)_{21}$ and $(Thr\text{-}Gly)_{22}$ were all poorer in temperature compensation, and this corresponds to what was previously found for the rare European Thr-Gly length alleles. Both lines of evidence therefore suggest an involvement of the $(Thr\text{-}Gly)_3$ repeat unit in the temperature compensation mechanism of *D.melanogaster*. The majority of Thr-Gly length variants appear to have an approximately a 24 h period at 29°C. If it is assumed that a 24h period is required for

fitness and is therefore less physiologically demanding than a shorter or longer clock, Australian populations are therefore particularly well adapted to their hotter climates.

The observation of a cline in Australia for the $(Thr\text{-}Gly)_{20}$ length allele, its frequency increasing with an increase in latitude, is evidence for natural selection. However, the lack of a cline for the other common allele the $(Thr\text{-}Gly)_{17}$ and therefore its relatively homogenous spatial differentiation could be due to its apparent ability to maintain its 24h period across temperature. Australia was presumably colonised from a mixture of African Asian, and European *D. melanogaster* populations by the advent of man (David and Capy, 1988) and variation can be accounted for by area effects, founder events, genetic drift, reduced gene flow, climatic and other local adaptations. Some differences in the characteristics of populations that occupy similar habitats and latitudes, but are widely separated geographically, may be accounted for by founder effects.. This is possibly the reason for the differences seen for the different Thr-Gly lengths from European and Australian populations.

Although the different Thr-Gly length variants of the Australian populations tested for their freerunning locomotor activity were not found to be completely consistent with the results described for the European Thr-Gly length variants, evidence was found in favour of the involvement of the $(Thr\text{-}Gly)_3$ repeat unit in temperature compensation. This again indicates the importance of the conformation and structure of the Thr-Gly repeat region in the mechanism of temperature compensation of the biological clock. Also, and perhaps more significantly, differential temperature compensation between southern alleles of the common length variants, $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$, between 25°C and 29°C was observed, which may go some way to interpreting the Australian cline, and which gave a consistent pattern with the European data. The distinct differences seen between northern and southern populations (Figs. 7.11a, b and 7.12a, b) suggests the involvement of another *X*-linked gene, since all the flies analysed had an identical genetic background aside from different *X* chromosomes. I will return to this point and its implications in the final chapter.

CHAPTER 8

8.1 INTRODUCTION

Many factors could be responsible for the presence of the European and Australian clines in the Thr-Gly alleles in *D. melanogaster* (Costa *et al.*, 1992, chapter 4). Chapters 5, 6 and 7 have invoked various behavioural phenotypes which could contribute to these spatial patterns. The finding of a significant difference in temperature compensation between the Thr-Gly length variants (chapter 5) suggests that a likely selective component could be temperature. Generally, when pooling across European origins, the shorter the Thr-Gly length, the poorer the temperature compensation. In addition the $(Thr\text{-}Gly)_{17}$ allele has a period nearer to 24 h at 29°C (Figs. 5.8a and b) and predominantly less active during the hottest parts of the day (chapter 7), whereas the $(Thr\text{-}Gly)_{20}$ variant is more active at cooler temperatures.

Evidence that temperature plays a direct role upon selection for different Thr-Gly repeat variants in *D. melanogaster*, does not yet exist, although circumstantial evidence suggests its involvement. The aim of this investigation was to test the hypothesis that when a mixed natural population of *D. melanogaster* is subjected to cold temperatures, selection will favour the $(Thr\text{-}Gly)_{20}$ allele because it is more frequent in cooler climates (Costa *et al.*, 1992, Chapter 4), and also that the $(Thr\text{-}Gly)_{17}$ variant would be favoured under warmer conditions. The experiment was set up initially by myself and Professor Rodolfo Costa (University of Padua, Italy) and the majority of the molecular work was carried out by Anna Goostrey, as part of her undergraduate thesis under my supervision.

8.2 METHODS

The following experiment was initiated in 1993, by Rodolfo Costa at the University of Padua, Italy. A natural population of *D. melanogaster* flies from Conselve (latitude 45.23), Italy, were collected and 81 isofemale lines were set up from this sample. Four or five females were then taken from each of the 81 isofemale lines and placed into a population cage, thus giving a representative sample of the initial Conselve population within the cage. After at least ten days a sample of parental females plus some of the F_1 generation of flies were placed in microfuge tubes and frozen for later analysis (Initial 2). Vials containing eggs from the initial population cage were then used to set up six replicate populations. Three cages , A, B and C were subjected to warm temperatures of 28°C and three other cages, D, E and F, to a cooler temperature of 20°C. The cages were kept at a constant temperature in a LD 12:12 cycle and maintained for the following 8 months, after which time fly samples were collected and again frozen for gene frequency analysis. Initial 1 was a sampled population taken in 1993 from the original isofemale lines. DNA extraction, PCR and agarose gel electrophoresis was carried out as described in chapter 2. Characterisation of the Thr-Gly variants of males and homozygous females was achieved by comparing the PCR product against those of known length markers. For heterozygous females, an extra band corresponding to the heteroduplex also forms. Therefore when a heterozygous female was identified from an initial PCR screen, the amplification was repeated and run against a known mixture of two Thr-Gly length variants from the known markers. The following heteroduplexes were made : $(Thr\text{-}Gly)_{14/17}$,$(Thr\text{-}Gly)_{14/20}$, $(Thr\text{-}Gly)_{14/21}$, $(Thr\text{-}Gly)_{14/23}$, $(Thr\text{-}Gly)_{17/20}$, $(Thr\text{-}Gly)_{17/21}$, $(Thr\text{-}Gly)_{17/23}$, $(Thr\text{-}Gly)_{20/21}$, $(Thr\text{-}Gly)_{21/23}$ by co-amplification of the appropriate DNA fragments. Figure 8.1 shows examples of the heteroduplex analysis.

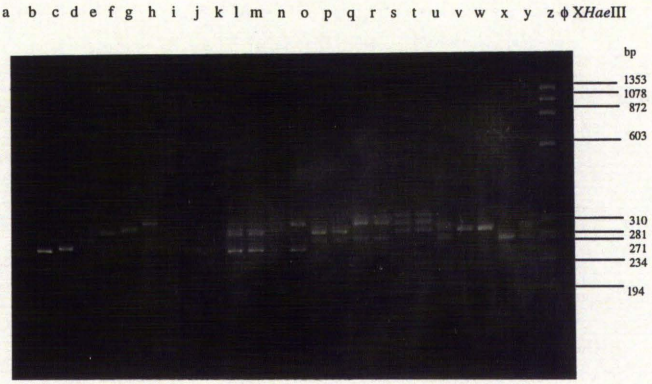


Figure 8.1 PCR-amplified *D. melanogaster* Thr-Gly length variants separated on a 3.5% low melting point gel (Nusieve). Amplification is with *5'per8* and *3'per2* , see Table 2.1 and Fig. 2.2. Lane a and y are φ X174 RF/*Hae*III size markers; lane b, $(Thr\text{-}Gly)_{14}$; lane c, $(Thr\text{-}Gly)_{17\text{-}2}$; lane d, $(Thr\text{-}Gly)_{17}$; lane e, $(Thr\text{-}Gly)_{20}$; lane f, $(Thr\text{-}Gly)_{21}$; lane g, $(Thr\text{-}Gly)_{23}$; lane h, blank; lane i and j, $(Thr\text{-}Gly)_{14/17}$; lane k and l, $(Thr\text{-}Gly)_{14/20}$; lane m and n, $(Thr\text{-}Gly)_{14/23}$; lane o and p, $(Thr\text{-}Gly)_{17/20}$; lane q and r, $(Thr\text{-}Gly)_{17/23}$; lane s and t, $(Thr\text{-}Gly)_{20/23}$; lanes u-z are flies used in the Conselve population analysis.

8.3 RESULTS

For all European populations so far investigated the majority of the Thr-Gly alleles observed were $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ (Table 8.1). Since this population was collected from southern Europe it also showed the characteristic higher proportion of $(Thr\text{-}Gly)_{17}$ alleles compared to $(Thr\text{-}Gly)_{20}$. Using the data from Table 8.1 a histogram was drawn to compare the final frequencies of the two most common length alleles (Fig 8.2) with the two initial population samples. It can be seen from Fig 8.2 that the Initial 2 population is a fair representation of the distribution of allele frequencies within the natural population of *D. melanogaster* in Conselve, Initial 1. A χ^2 test was performed to confirm this and the expected frequencies were based on the null hypothesis of no significant difference between the two populations. At the 5% level with df = 2, the χ^2 value obtained (0.33) did not represent a significant deviation from the null hypothesis. This meant there was no significant difference in allele frequency distribution between the population sample analysed previously (Initial 1) with the sample of flies the population cages were set up with (Initial 2). When comparing the final frequencies between the two types of population cages at 20°C and 28°C with the Initial 2 population a second χ^2 test was performed. The null hypothesis was that no significant difference will be observed between the frequency of alleles found within the three sets of populations (including Initial 2) maintained at the different temperatures. Therefore at the 5% level with df = 4, the value obtained for the χ^2 (3.142) does not reach significance, which suggests that the original allele frequencies found within Initial 2 and those found at the end of the experiment from either temperature do not differ significantly.

TABLE 8.1

Population	$(Thr\text{-}Gly)_{14}$	$(Thr\text{-}Gly)_{15}$	$(Thr\text{-}Gly)_{17}$	$(Thr\text{-}Gly)_{18}$	$(Thr\text{-}Gly)_{20}$	$(Thr\text{-}Gly)_{21}$	$(Thr\text{-}Gly)_{23}$	Total no. of alleles
A@ 28°C	0.013	0	0.481	0.078	0.377	0	0.052	77
B@ 28°C	0.014	0	0.592	0	0.394	0	0	71
C@ 28°C	0.026	0	0.494	0	0.312	0.026	0.143	77
average@ 28°C	0.018	0	0.52	0.027	0.36	0.001	0.067	
D@ 20°C	0	0.016	0.625	0	0.344	0	0.016	64
E@ 20°C	0.026	0	0.553	0	0.408	0	0.013	76
F@ 20°C	0.067	0	0.533	0	0.343	0	0.057	105
average@ 20°C	0.037	0.001	0.57	0	0.363	0	0.033	
Initial 1	0.036	0	0.58	0	0.321	0.018	0.045	112
Initial 2	0.05	0	0.55	0	0.35	0.027	0.027	183
Total no of alleles	26	1	420	6	270	9	33	765

Table 8.1 Frequencies of the different Thr-Gly length alleles in population cages at the end and at the beginning of the experiment.

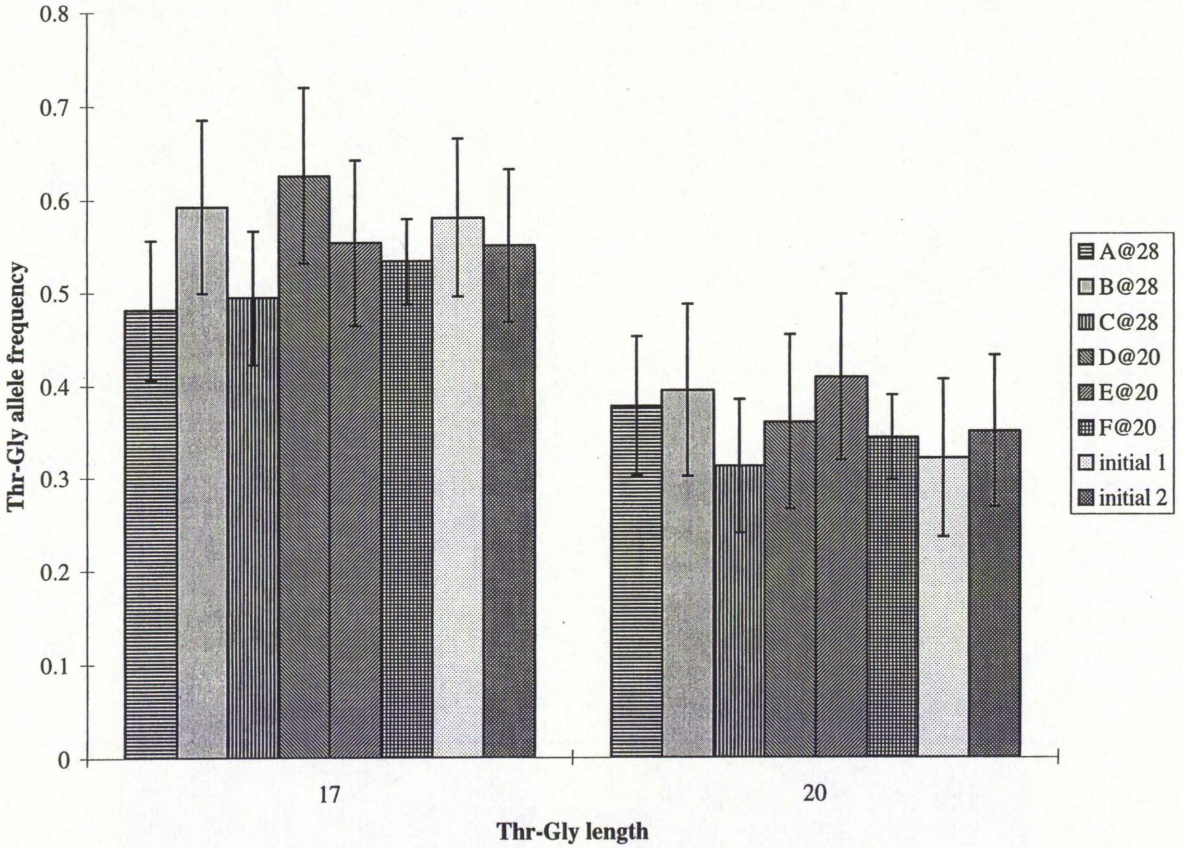


Figure 8.2 The common Thr-Gly allele frequencies for the different population cages (mean ± SEM)

8.4 DISCUSSION

The experiment described in this chapter was set up to test a simple thermal selection hypothesis. The initial population sample was taken from Conselve, where an average yearly temperature range of approximately 1°C to 27°C is found (Hulme *et al.*, 1995, Climates Impacts LINK Project). At locations with such temperatures predominance of the $(Thr\text{-}Gly)_{17}$ alleles might be expected (Costa *et al.*, 1992). Populations samples from Conselve maintained at the higher temperature 28°C (cages A, B and C) would be expected to show little or no change from the initial sample allele frequencies. Within populations maintained at the cooler 20°C (cages D, E and F), however, an increase in the frequency of the $(Thr\text{-}Gly)_{20}$ length variant and a decrease in the $(Thr\text{-}Gly)_{17}$ variant would be expected, under the simple thermal hypothesis proposed. Nevertheless, after the experiment had run its course, the populations of *D. melanogaster* maintained at the different temperatures showed no significant differences from one another in terms of their Thr-Gly allele frequencies (Table 8.1, Fig. 8.2). However, the constant temperature conditions used in the experiment were not a fair representation of the natural environment where wide diurnal fluctuations are generally found. The need for good temperature compensation may only arise under constantly fluctuating conditions and this could be a reason for the absence of any thermal selection. Also, eight months may not have been a sufficient length of time in which to detect a change in allele frequencies. Warm temperatures of 29°C permit 15-16 generations to pass and under cooler conditions, only 10-11 generations will occur within eight months. Furthermore, the temperature differences used may not have been extreme enough to represent the natural environment. Lowering the temperature however, would have caused the generation time to be too slow for reasonable experimental times and higher temperatures would have caused an increase in male sterility.

Powell (1971) and McDonald and Ayala (1974) have shown that selection operates on many enzyme polymorphisms in cage populations of *Drosophila*. Oakeshott (1979) scored gene frequencies at nine polymorphic loci twenty generations after population cage experiments of *D. melanogaster* were founded. For *Adh*, *Amylase* ($Amy^{4.6}$) and *Isocitrate dehydrogenase* ($Idh\text{-}NADP^{ss}$) gene frequencies, significant effects of temperature were observed. However, these relationships varied between loci and with environmental variability (Oakeshott, 1979). Whether the temperature-dependence is relevant to selection

in wild populations, however, remains doubtful, and in addition, the *in vitro* thermostabilities of these variants appear to be unrelated to the latitudinal clines and climatic analyses of the relevant enzymes (Wilks *et al.*, 1980; Oakeshott *et al.*, 1981).

A previous experiment, similar to that analysed here, was performed (R. Costa pers. comm.), where laboratory strains of *D. melanogaster* were used instead of the natural populations from Conselve. Cages were set up at the same two different temperatures with equal numbers of $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length variants. In this particular instance, it was observed that the $(Thr\text{-}Gly)_{20}$ length variants completely replaced the $(Thr\text{-}Gly)_{17}$ length variants, at both the warm and cool temperatures. However the two genotypes were not congenic and residual variation linked to *per* may have been responsible for this result, whereas in the experiment described in this chapter the flies carrying the different Thr-Gly variants were genetically heterogenous but largely congenic.

Although it has previously been shown that temperature is important for the spatial distribution of the flies (chapter 4, 5, 6 and 7), it is possible that the selection upon the Thr-Gly region is due to a combination of environmental factors which vary with latitude (Costa *et al.*, 1992). For example differences in photoperiod and light intensity occur with respect to latitude, and it has been shown that photoperiod and light wavelength have significant effects upon the circadian rhythms and growth characteristics of plants (Millar *et al.*, 1995b). The effect of photoperiod has yet to be investigated with the different Thr-Gly alleles of *D. melanogaster* .

In conclusion no putative single selective agent acting upon the Thr-Gly region was identified in these population cages, but more sophisticated population analyses with different interacting parameters, as suggested above, should be attempted in the future.

CHAPTER 9

In this thesis I have presented an analysis of the natural variation of the Thr-Gly region within the *per* gene in *D. melanogaster*, primarily from a phenotypic and population genetics perspective. This investigation began with the DNA sequence examination of the Thr-Gly repeat and its proximal flanking sequences in natural populations of *D. melanogaster* (chapter 3). Populations from Europe, north Africa, central Africa and Australia were sequenced and compared to those already published. The model previously proposed for the ancestral states of the Thr-Gly repeat being the $(Thr\text{-}Gly)_{20a}$ and $(Thr\text{-}Gly)_{23b}$ alleles (Peixoto *et al.*, 1993b; Rosato *et al.*, 1996) is supported by the sequence analysis presented in chapter 3. All alleles, including the new Thr-Gly length variants and new isolength alleles found, could be simply derived from either one or both of these by mechanisms such as slippage, gene conversion and base substitutions (Fig. 3.6). The alleles sequenced from central Africa in chapter 3 and others sequenced since (C. Pasetto pers. comm.), reveal some interesting observations. First, nothing less than 18 Thr-Gly pairs has been found on this continent, but the absence of shorter alleles contrasted with the presence of longer length variants, the $(Thr\text{-}Gly)_{20}$, $(Thr\text{-}Gly)_{21}$, $(Thr\text{-}Gly)_{22}$, $(Thr\text{-}Gly)_{23}$ and $(Thr\text{-}Gly)_{24}$. Clearly a reduction of variability has occurred during the colonisation of other continents by this species as described by Begun and Aquadro (1993), since not all of the African length variants have been found within European and Australian populations. Also within the African populations typed (C. Pasetto, pers. comm.), there was a high frequency of $(Thr\text{-}Gly)_{23}$ and $(Thr\text{-}Gly)_{20}$ variants, again suggesting that these are the ancestral alleles.

In *D. melanogaster*, the European and north African latitudinal cline observed for Thr-Gly length (Costa *et al.*, 1992), suggested that temperature-related selection is acting on the polymorphism. In order to demonstrate this, a number of different approaches can be considered. The classical way to 'prove' that natural selection is causing a clinal geographic distribution is to show that similar clines are present in other continents, as was shown in the *Adh* polymorphism in *D. melanogaster* (Oakeshott *et al.*, 1982; David *et al.*, 1989). Therefore, I carried out a similar study using natural populations of *D. melanogaster* from Australia (chapter 4; James *et al.*, 1995). Using a variety of spatial statistical tests it was suggested that the $(Thr\text{-}Gly)_{20}$ length variant was geographically structured along the eastern coast of Australia, since higher frequencies of this allele were found in the south than in the north (Fig. 4.3b). Also weak correlations in the same

direction were found for the $(Thr\text{-}Gly)_{23}$ allele (Fig. 4.3c). However, no cline was found for the $(Thr\text{-}Gly)_{17}$ length allele (Fig. 4.3a), although its frequencies across localities were higher in the north than those of the $(Thr\text{-}Gly)_{20}$ variant. This suggests one of two things, firstly selection is maintaining only the cline of the $(Thr\text{-}Gly)_{20}$ allele, and the European cline found for the $(Thr\text{-}Gly)_{17}$ is due to migration and population admixture, or secondly, there has not been a sufficient amount of time since the colonisation of Australia by *D. melanogaster* to allow the weak selection, thought to be acting on the Thr-Gly repeat, to be visible (Rosato *et al.*, ms. submitted). Also it should be noted that in Europe, the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles make up almost 90% of the genetic variability in each population studied. Consequently there is almost a statistical interdependence of the frequency of one allele upon the other since the other length variants are not present at sufficient enough frequencies to contribute. Thus a cline in one, may automatically mean a reciprocal cline in the other in Europe. In Australia (Table 4.5) the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ make up 83% the genetic variability, and perhaps the cline in the $(Thr\text{-}Gly)_{20}$ allele frequencies has not automatically led to a cline in the $(Thr\text{-}Gly)_{17}$ variant.

In addition various correlations with temperature were performed on revisiting the European cline and a significant correlation was found for temperature versus mean Thr-Gly length (Fig. 4.2), indicating an effect of this environmental variable in determining the cline. Also it was suggested from positive correlations that the $(Thr\text{-}Gly)_{17}$ allele is particularly adapted to warmer climates (Table 4.3). It was surprising however that the European $(Thr\text{-}Gly)_{20}$ frequency did not show a significant temperature correlation as they were spatially differentiated, and latitude correlates with temperature. Perhaps the European $(Thr\text{-}Gly)_{20}$ allele's tolerance to thermal challenge (chapter 5) can be used to explain this result, or other environmental factors are involved in the $(Thr\text{-}Gly)_{20}$ allele cline in Europe. In contrast, the Australian $(Thr\text{-}Gly)_{20}$ allele shows a significant negative correlation with mean yearly temperature (Table 4.8), indicating that this cline is also temperature related. The $(Thr\text{-}Gly)_{17}$ allele is generally homogenous across the eastern Australian populations studied, and since its cline in Europe was significantly stronger than that found for the $(Thr\text{-}Gly)_{20}$ allele (Costa *et al.*, 1992), it is again surprising that no reciprocal $(Thr\text{-}Gly)_{17}$ Australian cline has been observed. Perhaps the fact that the temperatures found in Australia are significantly elevated compared to those found in Europe (Gentilli, 1971), may contribute to the $(Thr\text{-}Gly)_{17}$ homogenous spatial structure.

Unequivocal evidence for selection gradients maintaining a cline would lie in the observation of parallel spatial patterns for the Thr-Gly length polymorphism in *D.*

melanogaster's sibling species *D. simulans*. However, no cline was found in *D. simulans*, and this species also showed a limited amount of polymorphism in the Thr-Gly region, which was in complete linkage disequilibrium with flanking haplotypes (Rosato *et al.*, 1994). The effectiveness of any selection for producing a cline depends upon the product of effective population size and selection coefficient, $N_e s$ (Kimura 1983 and see below). Since *D. simulans* has an effective population size three- to sixfold larger than *D. melanogaster* (Aquadro, 1992), this is consistent with a reduced effect of natural selection, where 'nearly neutral' mutations will be either eliminated or fixed.

Chapter 4 therefore appeared to show that perhaps selection in concert with historical factors and population constraints has produced these continental spatial structures of Thr-Gly polymorphism.

Chapter 5 examined any phenotypic difference in the freerunning locomotor activity periods of the Thr-Gly length variants at different temperatures. A global analysis revealed that the $(Thr\text{-}Gly)_{20}$ allele was the most efficient at maintaining its period between temperature (Figs. 5.8a and b). Genotypic differences were observed between spectrally derived periods of the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ alleles (Figs 5.3a and b), and this effect was also seen between the $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ alleles, where the shorter allele was more sensitive to temperature than the longer variant (Figs 5.4a and b). Although the division of lines between northern and southern European origin was somewhat arbitrary, differences between southern and northern alleles were observed. It was interesting to see the southern $(Thr\text{-}Gly)_{17}$ allele being less temperature sensitive in behaviour than the northern $(Thr\text{-}Gly)_{17}$ variant and the southern $(Thr\text{-}Gly)_{20}$ carrier. This may indicate a phenotypic advantage of the $(Thr\text{-}Gly)_{17}$ in its southern environment, as it is able to compensate for the higher, more physiologically stressful temperatures. Also the $(Thr\text{-}Gly)_{17}$ allele has a precise 24 h period at warmer temperatures (Fig. 5.8a and b), and perhaps this also helps to predispose it to southern regions of Europe because of any related fitness advantage accruing to this 'resonance' with the natural circadian period. In fact, the $(Thr\text{-}Gly)_{17}$ variant at 29°C and 18°C shows periods rather closer to 24 h than the $(Thr\text{-}Gly)_{20}$ allele (Fig. 5.8a and b). However the poorer temperature compensation of the $(Thr\text{-}Gly)_{17}$ carrier means that at temperatures lower than 18°C , the $(Thr\text{-}Gly)_{17}$ period could be further from 24 h than that of the $(Thr\text{-}Gly)_{20}$ allele (Fig. 5.8b). Consequently an overall explanation of the clinal patterns in Europe would be that the $(Thr\text{-}Gly)_{17}$ allele will generally have a higher frequency than the $(Thr\text{-}Gly)_{20}$ variant as its period is closer to 24 h.

However at colder temperatures, the poorer temperature compensation of the $(Thr\text{-}Gly)_{17}$ carrier means that its period will fall further from 24 h than that of the $(Thr\text{-}Gly)_{20}$ allele and as a result, the advantage rests with the $(Thr\text{-}Gly)_{20}$ variant in these conditions. This simple hypothesis explains the generally higher $(Thr\text{-}Gly)_{17}$ allele frequencies in Europe (Costa *et al.*, 1992). This kind of balancing selection appears to be a reasonable candidate to explain the spatial patterns observed. I did not have time to discover whether heterozygote advantage in female $(Thr\text{-}Gly)_{17}/(Thr\text{-}Gly)_{20}$ carriers might also contribute to the polymorphism.

As no flanking haplotypes have been found in association with the different Thr-Gly variants (Rosato *et al.*, 1996), this indicates that it is the Thr-Gly region itself which may be under selection. This is supported by *per in vitro* deletions of the Thr-Gly region and subsequent locomotor activity analyses (Fig. 5.16), which confirm the phenotypic differences seen in the naturally occurring alleles (J. M. Hennessy pers. comm.).

In chapter 7 the global analyses of the Australian Thr-Gly variants' freerunning locomotor activity showed the $(Thr\text{-}Gly)_{20}$ allele to be less efficient at maintaining its period across temperature than the $(Thr\text{-}Gly)_{17}$ variant (Fig 7.6a, b, Tables 7.1 and 7.2). The simple hypothesis proposed previously for the clinal distribution of length alleles in Europe due to the $(Thr\text{-}Gly)_{20}$ allele having a period nearer to 24 h in the cold, does not appear consistent for the Australian Thr-Gly length alleles since the $(Thr\text{-}Gly)_{17}$ variant has a period closer to 24 h at all temperatures when compared to the $(Thr\text{-}Gly)_{20}$ (Figs. 7.6a and b). In addition no significant correlations were found for temperature compensation and Thr-Gly length (Figs. 7.8a, b, c and 7.9a, b, c). However, when dissected further at a northern/southern level, it was found that the southern $(Thr\text{-}Gly)_{17}$ variant exhibited poorer temperature compensation between 25°C and 29°C than the $(Thr\text{-}Gly)_{20}$ allele (Figs. 7.11b and 12b). Furthermore, the northern Thr-Gly variants have appeared to have evolved a very robust temperature compensation mechanism (Figs. 7.11a and 7.12a). The dramatic differences observed between the extreme northern and southern Australian populations indicates the involvement of another *X*-linked gene or genes interacting with *per* because the autosomal genetic background in these lines was homogenised by using the attached-*X* lines.

Similar evidence was found for the involvement of the $(Thr\text{-}Gly)_3$ repeat unit in temperature compensation for both European and Australian Thr-Gly alleles, since the rarer variants that were out of phase with the 14-17-20-23 series were poorer in temperature

compensation (Fig. 5.9a and b; Figs. 7.7a-c and 7.8a-c). This supported the structural evidence that the $(Thr\text{-}Gly)_3$ is the conformational monomer (Castiglione-Morrelli *et al.*, 1995).

Chapter 6 described the phenotypic differences in the Thr-Gly variants energy expenditure by observing the LD activity profiles, which were revealed to differ subtly for one population studied, but was not replicated in another Italian population, although some results were similar (Appendix 6.11 and 6.12) . It was initially observed that the Conselve $(Thr\text{-}Gly)_{17}$ allele was relatively less active between ZT0-4 than the $(Thr\text{-}Gly)_{20}$ allele (Fig. 6.2), although this was not observed for the corresponding Lecce alleles (Fig. 6.4). This suggested a 'siesta' like effect for the $(Thr\text{-}Gly)_{17}$ allele which could have adaptive advantage in warmer southern European climates. These suggestive results were confirmed by the work of my colleague J. M. Hennessy working with Thr-Gly deletions in *D. melanogaster* transformants. $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_1$ transformants showed a significant reduction in the amount of activity in LD cycles compared to $(Thr\text{-}Gly)_{20}$ transformants between ZT2-7 at 18°C and this was extended to ZT10 at 29°C (J. M. Hennessy, pers. comm.). Similarly, the Yu *et al.*(1987b) Δ(Thr-Gly) transformant reveals an even more bimodal pattern of activity (J. M. Hennessy, pers. comm.). Clearly this 'siesta' is real and perhaps my observation on natural variants has detected this effect. Furthermore, the $(Thr\text{-}Gly)_{20}$ allele in both populations appeared to have significantly increased activity during ZT0-4 at 18°C, suggesting an adaptation to cooler climates, which may contibute to the European and Australian clines in this variant.

It is rare indeed that natural variation in a complex behavioural phenotype can be demonstrated to be under a single gene control, let alone be influenced by a small repetitive region within that gene, or be shown to involve selective processes. Under the 'nearly neutral theory' (Ohta, 1996), for selection to operate, the selection coefficient must be larger than the reciprocal of N_e, the effective population size. The common estimate of the evolutionary population size of *D. melanogaster*, N_e, is 10^6 (Kreitman 1996). Thus a selection coefficient slightly higher than 0.00001 would be visible to selection in *D. melanogaster*. What such a tiny coefficient would mean in terms of temperature compensation is difficult to imagine. Thus the small differences observed in temperature compensation between the different Thr-Gly length variants are potentially very large if viewed as selection coefficients, and are almost certainly visible to selection. Therefore, unless selection is very strong, laboratory experiments are usually orders of magnitude too

insensitive to reveal small, but evolutionary relevant selection coefficients among naturally occurring variants. This was shown to be the case in the experiments reported in chapter 8, which attempted to mimic the thermal selection previously found for the $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles in population cages. No effects were seen, underscoring the difficulties of detecting small selection coefficients in the laboratory. Also selection upon the Thr-Gly region could be due to a combination of latitudinally diverse environmental factors, including for example photoperiod and light intensity, and these have yet to be investigated with the different Thr-Gly alleles of *D. melanogaster* either in behavioural or population cage experiments.

The observation of a polymorphism affecting a behavioural phenotype is not unique. It has recently been shown that a dopamine D4 receptor (D4DR) exon III polymorphism is associated with the human personality trait of novelty seeking (Ebstein *et al.*, 1996; Benjamin *et al.*, 1996), specifically, longer repeat alleles show higher than average novelty seeking test scores. This corroboration between the two reports was especially impressive since a different personality questionnaire was employed, and an ethnically distinct population was examined. The association between the polymorphism and behaviour was evident within a family and between unrelated individuals, thereby showing the first replicated correlation between a specific genetic locus involved in neurotransmission and a complex behavioural trait. Interestingly it is thought that the different exon III length polymorphisms have different ligand binding properties (Van Tol *et al.*, 1992; Ashgari *et al.*, 1994). This set of studies takes the candidate gene approach: a gene which may be considered important in a phenotype is examined for natural variation that might explain phenotypic variation.

My work can be viewed in a similar way, with *per* being the candidate gene. This general approach has recently been applied to non-behavioural phenotypes. Natural populations of *D. melanogaster* harbouring polymorphic *Ultrabithorax* (*Ubx*) genes (Gibson and Hogness, 1996) appeared to show significant phenotypic consequences in the developmental stability of the third thoracic segment. This variation maps to the obvious candidate gene *Ubx*. Single-gene responses to strong selection for insecticide resistance have also been observed with single amino acid substitutions in a GABA receptor gene in *D. melanogaster* (MaKenzie and Batterham, 1994). Even more remarkable is that the same point mutation confers resistance in the natural populations of a wide range of insect species and so this study provides additional evidence for selection on a major gene having important evolutionary consequences. Interestingly , Endler (1986) states that such strong

selection could be common in nature. On the other hand, weak selection, for example, as has also been inferred for synonymous codon usage (Akashi, 1995, 1994), suggests even extremely minuscule fitness differences are being monitored by natural selection (Palopoli and Patel, 1996).

According to the long established neo-Darwinian view of evolutionary processes, large differences in phenotype are brought about by numerous gene substitutions, each of which have a relatively small effect on the phenotype individually (Fisher, 1930; Dobzhansky, 1937; Muller, 1940). On the basis of this micromutationist model the effects presented here by the Thr-Gly length polymorphisms in both DD and LD cycles could be adaptive but only in conjunction with other allelic substitutions at many loci. However, for this to apply no single substitution constitutes a major portion of the adaptation. On the other hand the phenotypic differences I have observed could be thought of as being caused by macromutations, since they are major enough to affect the geographical distributions of the polymorphisms. Macromutations are thought not to produce perfect adaptations by themselves either, but they are responsible for a large element of the adaptation. Clearly the Thr-Gly alleles appear to fall into the latter macromutational category as do the *Ubx* and *D4DR* alleles mentioned earlier. Kimura (1983) argued that it would be the mutations of intermediate value that will end up as the most frequent components of adaptations. However Hill (1982) proposed that each large effect mutation would contribute more to the selected trait than each of the small effect mutations, so that even if the former type are fixed less frequently, they will still contribute most of the response to selection. Lande (1983) concluded when considering the simultaneous segregation of micro- and macromutations on the evolution of quantitative traits, that unless selection is very strong and persistent, adaptation will usually result from substitutions of micromutations. The selection observed for the common Thr-Gly length alleles (Rosato *et al.*, ms. submitted), the $(Thr\text{-}Gly)_{17}$, $(Thr\text{-}Gly)_{20}$ and $(Thr\text{-}Gly)_{23}$ variants, is obviously strong enough to be visible and sufficiently persistent enough to maintain latitudinal clines in two different continents. In addition, adaptations in outbreeding populations must be usually based on genes with a conspicuous heterozygous effect, because otherwise they would be invisible to natural selection when they first arise (Charlesworth, 1990). Unfortunately heterozygous Thr-Gly length carriers have yet to be investigated.

In conclusion this study has revealed an insight into the molecular control of complex behaviour. To my knowledge this represents the first study of a behavioural gene where a correlation between the structure of a polymorphic region and the phenotype has

lent itself readily to a selective interpretation of continental spatial patterns. Further work to continue this theme would include:

i) studying the locomotor activity in both DD and LD of various heterozygous combinations of Thr-Gly alleles in females. Is there evidence for heterosis in temperature related behaviours ?

ii) to use the luciferase reporter genes (Brandes *et al.*, 1996) on different Thr-Gly backgrounds in order to examine the freerunning circadian periods of these variants at temperatures lower than 18°C. Would the hypothesis invoked previously to explain the clinal patterns based on the poorer temperature compensation of the $(Thr\text{-}Gly)_{17}$ allele and its corresponding significant departure from a 24 h period at colder temperatures compared to the $(Thr\text{-}Gly)_{20}$ be supported ?

iii) do the Thr-Gly variants show differences in behaviour under alternative photoperiodic regimes that might be relevant to natural populations ?

iv) is the molecular model described in chapter 5 involving Thr-Gly - PAS interactions valid ? Do the different Thr-Gly variants show genotype specific interactions with the upstream *per* domain ?

Clearly these important questions need to be examined. My work has generated several interesting observations and has perhaps finally found a role for the enigmatic Thr-Gly region.

APPENDICES FOR CHAPTER 4

Appendix 4.1 Computation of the spatial autocorrelation coefficient, Morans *I*

* As an example I have taken the four Australian localities MO, BRFS, K, and AG (Fig. 4.1, Table 4.5, Table A) and calculated the smallest distance class (0, 432.2 Km) for the frequencies of the $(Thr\text{-}Gly)_{17}$ allele, q_i, at the *i*th locality. The mean frequency of the $(Thr\text{-}Gly)_{17}$ allele across these four localities is q.

TABLE A

i, locality	q_i, $(Thr\text{-}Gly)_{17}$ frequency	$q_i - q$
MO	0.43	-0.0382
BRFS	0.38	-0.0925
K	0.52	0.0475
AG	0.56	0.085

* There are 2 pairs of localities, BRFS-K, and K-AG, within the smallest distance class. The formula employed is:

$$I = N\Sigma\Sigma w_{ij}(q_i - q)(q_j - q) / W\Sigma(q_i - q)^2 ,$$

where q_i and q_j are the frequencies of the allele of interest at the *i*th and *j*th locality, q is their mean across the N (4 in this case) localities, w_{ij} is equal to 1 for all the pairs of localities falling in the distance class studied and equal to 0 for all other pairs, and W is the sum of all w_{ij} values in that distance class.

* To compute the expression it is broken down into parts. For the numerator all products of pairs of localities are needed, Table B

TABLE B

Locality, i	locality j	weight, w_{ij}	$(q_i - q)(q_j - q)$	$w_{ij}(q_i - q)(q_j - q)$
BRFS	K	1	(-0.0925)(0.0475)	-0.004394
K	AG	1	(0.0475)(0.0875)	0.004156
MO	BRFS	0	(-0.0382) -(0.0925)	0
:	:	:	:	:
:	:	:	:	:
MO	AG	0	(0.0382)(0.085)	0
		$\Sigma w_{ij} = W = 2$	$\Sigma\Sigma w_{ij}(q_i - q)(q_j - q)$	= -0.000238

* Since most weights are zero, very few multiplications are actually carried out. The weighted sum of the products

$$\Sigma\Sigma w_{ij}(q_i - q)(q_j - q) = -0.000238.$$

* The sum of the squares of the deviates from the mean, q, of all localities is

$$\Sigma (q_i - q)^2 = (-0.0382)^2 + (-0.0925)^2 + (0.0475)^2 + (0.0875)^2$$
$$= 0.0195$$

* Therefore $I = 4(-0.000238)/2(0.0195)$
$$= -0.0244$$

Appendix 4.2 Royaltey-Astrachan-Sokal non-parametric test of departure from randomness (Royaltey *et al.*, 1975)

* connect all localities by Delaunay network (Fig. 4.1), this is a triangulation of neighbouring localities based on criteria of geographic contiguity and permissible relationships between any pair or localities indicated by an edge connecting the pair (Brasseil and Reif, 1979).

* rank the frequencies (1-20) of allele under investigation, in this case the $(Thr\text{-}Gly)_{20}$, at each locality

* edge lengths, calculate the rank differences, regardless of sign, between the frequencies of pairs of connected localities, L. The test comprises of an examination of the distribution of these edge lengths. $\Sigma L = 260$.

* number of incidences (sum of connections (edges) for all localities (vertices, i))

$$\Sigma I = 92 = 2e$$

* observed mean edge length, where e is the total number of edges in the Delaunay network

$$\sum_{i=1}^{e} L_i / e = 260/46 = 5.652$$

* the edge connected by two points in the network can vary from 1 to n-1. The probability that the length of the edge is X when X is 1 to n-1. Therefore the expected number of edges with length X in the network is $2e(n-X)/n(n-1)$, where e is the total number of edges in the triangulation. The expected mean edge length is then

$$n+1/3 = 21/3 = 7$$

* correction for continuity

$$1/2e = 1/92 = 0.011$$

* the formulae for the variances of the means are functions of the indices, I, for any of the vertices in the network. The variance of the total edge length of the connections is

$$\delta^2 (\sum_{i=1}^{e} L_i / e) = (n+1)[(n-4)\,\Sigma I^2 - (\Sigma I)^2 + 4(n-1)]180$$

* from the variance of the total edge length the standard deviation of mean edge length can be obtained

$$\delta(L) = \delta (\sum_{i=1}^{e} L_i / e)/e^2 = \sqrt{(n+1)\,[(n-4)\,\Sigma I^2 - (\Sigma I)^2 + 4(n-1)\,\Sigma I]\,/45(\Sigma I)^2}$$

$$\sqrt{21}[16 \,.\, 36 - 92^2 + 4 \,.\, 19 \,.\, 92]/45 \,.\, 92^2$$

$$= 0.524$$

* minimisation of the rank differences results in a low standard deviation, which is indicative of the smoothness of transition between localities across the network

* the observed mean edge length was compared with the expected mean edge length using a student's *t* test. the *t*-test assumes a fairly normal distribution of mean edge lengths

t_s = (observed mean edge length-expected mean edge length ± correction for continuity)/(standard deviation of mean edge length)

t_s = 5.652 - 7 + 0.011/0.524

= -2.55 $P[\propto] < 0.02$

* the negative *t* value shows that the departure from randomness results from an underlying clinal pattern of $(Thr\text{-}Gly)_{20}$ allele frequencies.

APPENDICES FOR CHAPTER 5

Appendix 5.1. Autocorrelation, Planned comparisons between pooled Thr-Gly length alleles at differe temperatures. (*, p<0.05, MS error, 0.88, df, 1481)

TEMP /°C	Thr-Gly LENGTH	MEAN PERIOD/h	difference between means	MS	df	F	p
18	14	23.69	0.59	17.34	1	19.6	0.000*
29	14	24.28					
18	15	23.88	0.2	0.31	1	0.35	0.560
29	15	24.08					
18	17	23.62	0.44	24.84	1	28.10	0.000*
29	17	24.06					
18	20	23.68	0.21	4.24	1	4.80	0.029*
29	20	23.89					
18	21	23.75	0.14	0.66	1	0.74	0.390
29	21	23.89					
18	23	23.9	0.12	0.72	1	0.81	0.370
29	23	24.02					
18	24	23.17	0.30	0.71	1	0.80	0.370
29	24	23.47					

Appendix 5.2. Two way fixed ANOVA, Spectral (*, p<0.05)

SOURCE	MS	df	F	p
GENOTYPE(G)	2.4	6	3.79	0.001*
TEMPERATURE(T)	1.27	1	2.01	0.156
G*T	2.64	6	4.16	0.000*
ERROR	0.634	1481		

Appendix 5.3. Spectral, Planned comparisons between pooled Thr-Gly length alleles at different temperatures. (*, p<0.05, MS error, 0.634, df, 1481)

TEMP /°C	Thr-Gly LENGTH	MEAN PERIOD/h	difference between means	MS	df	F	p
18	14	23.79	0.24	2.86	1	4.51	0.034*
29	14	24.03					
18	15	23.92	0.27	0.54	1	0.85	0.360
29	15	23.65					
18	17	23.74	0.15	3.01	1	4.75	0.029*
29	17	23.89					
18	20	23.72	0.05	0.33	1	0.53	0.470
29	20	23.67					
18	21	23.69	0.23	1.68	1	2.66	0.100
29	21	23.92					
18	23	24.08	0.29	4.32	1	6.82	0.009*
29	23	23.79					
18	24	23.77	0.65	3.42	1	5.4	0.020*
29	24	23.12					

Appendix 5.4. Unnested ANOVA on $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles. (*, $p<0.05$)

SPECTRAL				
EFFECT	MS	df	F	p
GENOTYPE(G)	3.18	1	4.67	0.031*
TEMP(T)	0.48	1	0.71	0.399
T*G	2.47	1	3.62	0.057
ERROR	0.68	888	0.68	

Appendix 5.5. Unnested ANOVA on $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ length alleles. (*, $p<0.05$)

SPECTRAL				
EFFECT	MS	df	F	p
GENOTYPE(G)	0.83	1	1.46	0.228
TEMP(T)	0.10	1	0.17	0.679
T*G	3.88	1	6.81	0.009*
ERROR	0.569	371		

Appendix 5.6. Two way fixed ANOVA, Spectral (*, $p<0.05$)

SOURCE	MS	d.f	F	p
TEMPERATURE(T)	0.08	1	0.14	0.711
LINE(L)	3.89	37	7.05	0.000*
T*L	1.22	37	2.21	0.000*
ERROR	0.55	1419		

Appendix 5.7. Planned comparisons of nested $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles between different temperatures (*, $p<0.05$).

AUTOCORRELATION (MS error, 0.803; df, 831)

line	genotype	mean period @ 18°C	mean period @ 29°C	MS	df	F	p -level
CONS	17A	23.97	24.45	3.71	1	4.62	0.032*
CONS	17B	23.87	23.78	0.08	1	0.10	0.750
CONS	17C	23.32	24.25	9.26	1	11.53	0.001*
LE	17	23.26	23.83	2.06	1	2.56	0.110
PI.1	17	23.42	23.61	0.23	1	0.29	0.590
PI.9	17	23.60	24.25	3.06	1	3.81	0.051(*)
CAS	17	23.31	23.73	4.16	1	5.18	0.023*
RET.9	17	23.81	23.96	0.19	1	0.24	0.620
LEC60	17	23.32	24.30	9.44	1	11.76	0.001*
LEC12	17	24.14	24.25	0.14	1	0.18	0.670
NW6	17	23.89	23.95	0.37	1	0.04	0.840
COG	20	23.48	23.67	0.50	1	0.62	0.430
LE	20	23.83	23.92	0.06	1	0.08	0.780
PI.5	20	23.75	23.63	0.04	1	0.05	0.820
PI.8	20	23.85	24.13	0.34	1	0.42	0.520
CAS	20	23.79	23.40	1.44	1	1.80	0.180
RET.4	20	24.34	24.68	0.75	1	0.93	0.360
CONS	20A	23.42	23.72	1.20	1	1.50	0.220
CONS	20B	23.36	24.18	8.18	1	10.20	0.001*
LEC6	20	23.93	24.00	0.03	1	0.04	0.840
LEC3	20	23.46	24.25	4.05	1	5.05	0.030*
NW1	20	23.35	23.56	0.28	1	0.34	0.560

Appendix 5.8. Nested ANOVA on $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles (*, $p<0.05$)

SPECTRAL				
EFFECT	MS	df	F	p
GENOTYPE(G)	2.11	1	3.72	0.054
TEMP(T)	0.69	1	1.22	0.269
LINE(L)	5.14	20	9.07	0.000*
T*G	0.15	1	0.28	0.612
L*T	1.12	20	1.98	0.007*
ERROR	0.567	831		

Appendix 5.9. Planned comparisons of nested $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$ length alleles between different temperatures (*, $p<0.05$).

SPECTRAL (MS error, 0.567, df, 831)

line	genotype	mean period @ 18°C	mean period @ 29°C	MS	df	F	p -level
CONS	17A	23.94	24.55	5.81	1	10.25	0.001*
CONS	17B	23.31	23.71	1.60	1	2.83	0.093
CONS	17C	23.66	23.91	0.68	1	1.20	0.270
LE	17	23.49	23.78	0.53	1	0.94	0.330
PI.1	17	23.75	23.91	0.16	1	0.29	0.590
PI.9	17	23.99	23.81	0.24	1	0.42	0.520
CAS	17	23.21	23.09	0.31	1	0.55	0.460
RET.9	17	23.84	23.94	0.08	1	0.14	0.710
LEC60	17	24.18	24.31	0.18	1	0.32	0.570
LEC12	17	24.13	24.14	0.00	1	0.00	0.980
NW6	17	24.48	23.84	3.13	1	5.52	0.019*
COG	20	23.70	23.39	1.31	1	2.31	0.130
LE	20	23.81	23.91	0.73	1	0.13	0.720
PI.5	20	23.37	23.61	0.16	1	0.29	0.590
PI.8	20	24.01	24.28	0.33	1	0.58	0.450
CAS	20	23.69	23.12	3.11	1	5.50	0.019*
RET.4	20	23.81	24.20	1.00	1	1.76	0.184
CONS	20A	23.47	23.28	0.49	1	0.87	0.350
CONS	20B	23.72	23.79	0.06	1	0.11	0.750
LEC6	20	23.90	24.40	1.83	1	3.23	0.070
LEC3	20	23.88	24.27	0.98	1	1.73	0.190
NW1	20	23.73	23.20	1.75	1	3.08	0.079

Appendix 5.10. Nested ANOVA on $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$ length alleles. (*, $p<0.05$)

SPECTRAL				
EFFECT	MS	df	F	p
GENOTYPE(G)	1.95	1	3.74	0.054
TEMP(T)	0.31	1	0.60	0.440
LINE(L)	2.79	6	5.36	0.000*
T*G	1.52	1	2.91	0.089
L*T	0.94	6	1.80	0.099
ERROR	0.521	341		

Appendix 5.11. Autocorrelation, Spearman rank order correlations for period length of Thr-Gly variants At 18°C and 29°C(*, p<0.05)

	18°C			29°C		
Thr-Gly Variant	r_s	p	df	r_s	p	df
$(Thr\text{-}Gly)_{14}$	-0.01	0.898	101	0.10	0.346	99
$(Thr\text{-}Gly)_{17}$	0.07	0.263	264	0.08	0.188	244
$(Thr\text{-}Gly)_{20}$	-0.11	0.116	195	-0.17	0.018*	181
$(Thr\text{-}Gly)_{21}$	-0.01	0.932	67	-0.31	0.012*	62
$(Thr\text{-}Gly)_{23}$	0.09	0.390	88	0.37	0.001*	79

Appendix 5.12. Spectral, Spearmans rank order correlations for period length of Thr-Gly variants At 18°C and 29°C(*, p<0.05)

	18°C			29°C		
Thr-Gly Variant	r_s	p	df	r_s	p	df
$(Thr\text{-}Gly)_{14}$	0.05	0.593	101	0.19	0.059	99
$(Thr\text{-}Gly)_{17}$	0.19	0.002*	264	0.16	0.013*	244
$(Thr\text{-}Gly)_{20}$	-0.04	0.625	195	-0.20	0.006*	181
$(Thr\text{-}Gly)_{21}$	-0.16	0.200	67	-0.08	0.513	62
$(Thr\text{-}Gly)_{23}$	0.16	0.124	88	0.39	0.000*	79

APPENDICES FOR CHAPTER 6

Appendix 6.1 Planned comparisons between $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$

CONSELVE (MS error, 1681.6; df, 2952)

WINDOW	TEMP/ °C	MS	df	F	p -level
3-5pm ZT6-8	18	4432.63	1	2.64	0.105
5-7pm ZT8-10	18	4522.58	1	2.69	0.101
7-9pm ZT10-12	18	7070.93	1	4.20	0.040*
9-11pm ZT12-14	18	5296.91	1	3.15	0.076
11-1am ZT14-16	18	1021.63	1	0.61	0.436
1-3am ZT16-18	18	913.02	1	0.54	0.461
3-5am ZT18-20	18	1964.47	1	1.17	0.280
5-7am ZT20-22	18	2709.14	1	1.61	0.204
7-9am ZT22-0	18	14057.20	1	8.36	0.004*
9-11am ZT0-2	18	41780.41	1	24.85	0.000*
11-1pm ZT2-4	18	27492.03	1	16.35	0.000*
1-3pm ZT4-6	18	12625.05	1	7.51	0.006*
3-5pm ZT6-8	29	0.42	1	0.00	0.987
5-7pm ZT8-10	29	66.46	1	0.04	0.842
7-9pm ZT10-12	29	50.30	1	0.03	0.863
9-11pm ZT12-14	29	106.84	1	0.06	0.801
11-1am ZT14-16	29	847.64	1	0.50	0.478
1-3am ZT16-18	29	1347.47	1	0.80	0.371
3-5am ZT18-20	29	1635.53	1	0.97	0.324
5-7am ZT20-22	29	16424.51	1	9.77	0.002*
7-9am ZT22-0	29	58022.03	1	34.50	0.000*
9-11am ZT0-2	29	43671.14	1	25.97	0.000*
11-1pm ZT2-4	29	4968.30	1	2.95	0.086
1-3pm ZT4-6	29	0.74	1	0.00	0.983

Appendix 6.2 Planned comparisons between $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$

CONSELVE (MS error, 1681.6; df, 2952)

WINDOW	TEMP/ °C	MS	df	F	p -level
3-5pm ZT6-8	18	6132.67	1	3.65	0.056
5-7pm ZT8-10	18	3299.36	1	1.96	0.161
7-9pm ZT10-12	18	4283.80	1	2.55	0.111
9-11pm ZT12-14	18	0.29	1	0.00	0.989
11-1am ZT14-16	18	15.43	1	0.01	0.923
1-3am ZT16-18	18	18.87	1	0.01	0.916
3-5am ZT18-20	18	9.80	1	0.01	0.939
5-7am ZT20-22	18	174.14	1	0.10	0.748
7-9am ZT22-0	18	3603.07	1	2.14	0.143
9-11am ZT0-2	18	13900.75	1	8.27	0.004*
11-1pm ZT2-4	18	9714.93	1	5.78	0.016*
1-3pm ZT4-6	18	7605.16	1	4.52	0.034*
3-5pm ZT6-8	29	12.77	1	0.01	0.931
5-7pm ZT8-10	29	91.70	1	0.05	0.815
7-9pm ZT10-12	29	4415.72	1	2.63	0.105
9-11pm ZT12-14	29	136.47	1	0.08	0.776
11-1am ZT14-16	29	189.98	1	0.11	0.737
1-3am ZT16-18	29	453.46	1	0.27	0.604
3-5am ZT18-20	29	2633.29	1	1.57	0.211
5-7am ZT20-22	29	2884.11	1	1.72	0.190
7-9am ZT22-0	29	379.98	1	0.23	0.635
9-11am ZT0-2	29	5659.24	1	3.37	0.067
11-1pm ZT2-4	29	2162.55	1	1.29	0.257
1-3pm ZT4-6	29	9.61	1	0.01	0.940

Appendix 6.3 Planned comparisons between $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$

LECCE (MS error, 1154.89; df, 2316)

WINDOW	TEMP/ °C	MS	df	F	p -level
3-5pm ZT6-8	18	4195.10	1	3.63	0.057
5-7pm ZT8-10	18	11954.84	1	10.35	0.001*
7-9pm ZT10-12	18	2712.51	1	2.35	0.126
9-11pm ZT12-14	18	968.90	1	0.84	0.360
11-1am ZT14-16	18	21.18	1	0.02	0.892
1-3am ZT16-18	18	2.06	1	0.00	0.966
3-5am ZT18-20	18	65.05	1	0.06	0.812
5-7am ZT20-22	18	0.12	1	0.00	0.992
7-9am ZT22-0	18	342.69	1	0.30	0.586
9-11am ZT0-2	18	1.16	1	0.00	0.975
11-1pm ZT2-4	18	500.28	1	0.43	0.510
1-3pm ZT4-6	18	2286.98	1	1.98	0.160
3-5pm ZT6-8	29	2.74	1	0.00	0.961
5-7pm ZT8-10	29	335.02	1	0.29	0.590
7-9pm ZT10-12	29	2765.32	1	2.39	0.122
9-11pm ZT12-14	29	22692.16	1	19.65	0.000*
11-1am ZT14-16	29	1.87	1	0.00	0.968
1-3am ZT16-18	29	26.61	1	0.02	0.879
3-5am ZT18-20	29	0.27	1	0.00	0.988
5-7am ZT20-22	29	243.49	1	0.21	0.646
7-9am ZT22-0	29	407.81	1	0.35	0.552
9-11am ZT0-2	29	5706.47	1	4.94	0.026*
11-1pm ZT2-4	29	822.89	1	0.71	0.399
1-3pm ZT4-6	29	248.52	1	0.22	0.643

Appendix 6.4 Planned comparisons between $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$

LECCE (MS error, 1154.89; df, 2316)

WINDOW	TEMP/ ^{o}C	MS	df	F	p -level
3-5pm ZT6-8	18	1536.08	1	1.33	0.249
5-7pm ZT8-10	18	815.22	1	0.71	0.400
7-9pm ZT10-12	18	2147.84	1	1.86	0.173
9-11pm ZT12-14	18	179.50	1	0.16	0.693
11-1am ZT14-16	18	124.08	1	0.11	0.743
1-3am ZT16-18	18	38.71	1	0.03	0.855
3-5am ZT18-20	18	29.87	1	0.03	0.872
5-7am ZT20-22	18	3.22	1	0.00	0.958
7-9am ZT22-0	18	934.74	1	0.81	0.368
9-11am ZT0-2	18	8469.68	1	7.33	0.007*
11-1pm ZT2-4	18	909.34	1	0.79	0.375
1-3pm ZT4-6	18	1285.75	1	1.11	0.291
3-5pm ZT6-8	29	23.46	1	0.02	0.887
5-7pm ZT8-10	29	45.07	1	0.04	0.843
7-9pm ZT10-12	29	246.62	1	0.21	0.644
9-11pm ZT12-14	29	76.91	1	0.07	0.796
11-1am ZT14-16	29	943.95	1	0.82	0.366
1-3am ZT16-18	29	355.48	1	0.31	0.579
3-5am ZT18-20	29	401.01	1	0.35	0.556
5-7am ZT20-22	29	13.87	1	0.01	0.913
7-9am ZT22-0	29	3.17	1	0.00	0.958
9-11am ZT0-2	29	1351.44	1	1.17	0.279
11-1pm ZT2-4	29	1074.18	1	0.93	0.335
1-3pm ZT4-6	29	35.18	1	0.03	0.861

Appendix 6.5 Pearson product moment correlations for the actual activity of Thr-Gly variants at 18°C and 29°C in various 2h windows(*, p<0.05)

CONSELVE		18°C			29°C		
Window		*r*	p	df	*r*	p	df
3-5pm	ZT6-8	0.12	0.159	137	0.07	0.424	119
5-7pm	ZT8-10	0.04	0.655	137	-0.11	0.216	119
7-9pm	ZT10-12	0.05	0.571	137	0.07	0.480	119
9-11pm	ZT12-14	0.03	0.707	137	-0.03	0.746	119
11-1am	ZT14-16	0.08	0.362	137	-0.02	0.855	119
1-3am	ZT16-18	0.12	0.170	137	-0.04	0.660	119
3-5am	ZT18-20	0.03	0.703	137	-0.01	0.290	119
5-7am	ZT20-22	0.10	0.228	137	-0.04	0.642	119
7-9am	ZT22-0	0.19	0.023*	137	0.02	0.801	119
9-11am	ZT0-2	0.19	0.030*	137	0.18	0.050*	119
11-1pm	ZT2-4	0.25	0.004*	137	0.19	0.039*	119
1-3pm	ZT4-6	0.21	0.013*	137	0.06	0.541	119

Appendix 6.6 Pearson product moment correlations for the actual activity of Thr-Gly variants at 18°C and 29°C in various 2h windows(*, p<0.05)

LECCE		18°C			29°C		
Window		*r*	p	df	*r*	p	df
3-5pm	ZT6-8	0.15	0.125	110	0.05	0.618	91
5-7pm	ZT8-10	0.04	0.688	110	0.07	0.511	91
7-9pm	ZT10-12	-0.02	0.837	110	-0.09	0.396	91
9-11pm	ZT12-14	0.04	0.661	110	-0.06	0.594	91
11-1am	ZT14-16	0.11	0.269	110	0.21	0.045*	91
1-3am	ZT16-18	0.07	0.480	110	0.13	0.239	91
3-5am	ZT18-20	0.06	0.552	110	0.16	0.137	91
5-7am	ZT20-22	0.02	0.878	110	0.07	0.495	91
7-9am	ZT22-0	-0.09	0.327	110	0.04	0.745	91
9-11am	ZT0-2	-0.11	0.248	110	0.01	0.905	91
11-1pm	ZT2-4	-0.04	0.683	110	0.14	0.193	91
1-3pm	ZT4-6	0.16	0.092	110	-0.01	0.929	91

Appendix 6.7 Planned comparisons between $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$

CONSELVE (MS error, 0.008; df, 2952)

WINDOW	TEMP/ °C	MS	df	F	p -level
3-5pm ZT6-8	18	0.02	1	3.03	0.082
5-7pm ZT8-10	18	0.09	1	12.37	0.000*
7-9pm ZT10-12	18	0.21	1	28.54	0.000*
9-11pm ZT12-14	18	0.04	1	5.12	0.023*
11-1am ZT14-16	18	0.03	1	4.34	0.037*
1-3am ZT16-18	18	0.03	1	4.25	0.039*
3-5am ZT18-20	18	0.05	1	6.59	0.010*
5-7am ZT20-22	18	0.05	1	7.2	0.007*
7-9am ZT22-0	18	0.09	1	11.66	0.001*
9-11am ZT0-2	18	0.02	1	3.11	0.078
11-1pm ZT2-4	18	0.11	1	14.65	0.000*
1-3pm ZT4-6	18	0.02	1	2.59	0.108
3-5pm ZT6-8	29	0.00	1	0.15	0.746
5-7pm ZT8-10	29	0.00	1	0.50	0.822
7-9pm ZT10-12	29	0.02	1	2.55	0.111
9-11pm ZT12-14	29	0.15	1	19.53	0.000*
11-1am ZT14-16	29	0.03	1	3.49	0.062
1-3am ZT16-18	29	0.01	1	0.88	0.346
3-5am ZT18-20	29	0.00	1	0.05	0.825
5-7am ZT20-22	29	0.06	1	8.44	0.004*
7-9am ZT22-0	29	0.06	1	8.59	0.003*
9-11am ZT0-2	29	0.06	1	7.76	0.005*
11-1pm ZT2-4	29	0.03	1	4.49	0.034*
1-3pm ZT4-6	29	0.10	1	0.10	0.754

Appendix 6.8 Planned comparisons between $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$

CONSELVE (MS error, 0.008; df, 2952)

WINDOW	TEMP/ ^{o}C	MS	df	F	p -level
3-5pm ZT6-8	18	0.01	1	0.77	0.381
5-7pm ZT8-10	18	0.00	1	0.12	0.727
7-9pm ZT10-12	18	0.01	1	1.99	0.158
9-11pm ZT12-14	18	0.00	1	0.32	0.572
11-1am ZT14-16	18	0.00	1	0.47	0.493
1-3am ZT16-18	18	0.00	1	0.24	0.626
3-5am ZT18-20	18	0.00	1	0.23	0.632
5-7am ZT20-22	18	0.01	1	1.25	0.264
7-9am ZT22-0	18	0.03	1	3.73	0.053
9-11am ZT0-2	18	0.00	1	0.68	0.411
11-1pm ZT2-4	18	0.01	1	1.80	0.180
1-3pm ZT4-6	18	0.01	1	1.31	0.252
3-5pm ZT6-8	29	0.01	1	0.96	0.327
5-7pm ZT8-10	29	0.00	1	0.12	0.731
7-9pm ZT10-12	29	0.05	1	6.21	0.013*
9-11pm ZT12-14	29	0.05	1	6.59	0.010*
11-1am ZT14-16	29	0.02	1	3.29	0.070
1-3am ZT16-18	29	0.03	1	4.21	0.040*
3-5am ZT18-20	29	0.04	1	5.06	0.025*
5-7am ZT20-22	29	0.05	1	6.94	0.008*
7-9am ZT22-0	29	0.03	1	3.54	0.060*
9-11am ZT0-2	29	0.08	1	10.84	0.001*
11-1pm ZT2-4	29	0.07	1	9.83	0.002*
1-3pm ZT4-6	29	0.01	1	1.07	0.300

Appendix 6.9 Planned comparisons between $(Thr\text{-}Gly)_{17}$ and $(Thr\text{-}Gly)_{20}$

LECCE (MS error, 0.009; df, 2316)

WINDOW	TEMP/ °C	MS	df	F	p -level
3-5pm ZT6-8	18	0.03	1	2.85	0.091
5-7pm ZT8-10	18	0.09	1	9.45	0.002*
7-9pm ZT10-12	18	0.00	1	0.00	0.978
9-11pm ZT12-14	18	0.04	1	4.27	0.039*
11-1am ZT14-16	18	0.00	1	0.03	0.856
1-3am ZT16-18	18	0.00	1	0.36	0.546
3-5am ZT18-20	18	0.01	1	0.89	0.344
5-7am ZT20-22	18	0.00	1	0.17	0.683
7-9am ZT22-0	18	0.05	1	4.97	0.026*
9-11am ZT0-2	18	0.01	1	1.25	0.263
11-1pm ZT2-4	18	0.00	1	0.24	0.624
1-3pm ZT4-6	18	0.03	1	2.67	0.103
3-5pm ZT6-8	29	0.00	1	0.06	0.805
5-7pm ZT8-10	29	0.04	1	3.74	0.053
7-9pm ZT10-12	29	0.00	1	0.43	0.510
9-11pm ZT12-14	29	0.21	1	22.20	0.000*
11-1am ZT14-16	29	0.04	1	4.19	0.041*
1-3am ZT16-18	29	0.03	1	2.85	0.091
3-5am ZT18-20	29	0.03	1	3.28	0.070
5-7am ZT20-22	29	0.08	1	8.09	0.005*
7-9am ZT22-0	29	0.00	1	0.37	0.541
9-11am ZT0-2	29	0.07	1	7.56	0.006*
11-1pm ZT2-4	29	0.01	1	0.64	0.423
1-3pm ZT4-6	29	0.01	1	1.52	0.217

Appendix 6.10 Planned comparisons between $(Thr\text{-}Gly)_{14}$ and $(Thr\text{-}Gly)_{23}$

LECCE (MS error, 0.009; df, 2316)

WINDOW	TEMP/ ^{o}C	MS	df	F	p -level
3-5pm ZT6-8	18	0.04	1	4.26	0.039*
5-7pm ZT8-10	18	0.00	1	0.00	0.978
7-9pm ZT10-12	18	0.00	1	0.03	0.852
9-11pm ZT12-14	18	0.00	1	0.30	0.587
11-1am ZT14-16	18	0.01	1	0.54	0.462
1-3am ZT16-18	18	0.00	1	0.21	0.647
3-5am ZT18-20	18	0.00	1	0.00	0.989
5-7am ZT20-22	18	0.00	1	0.00	0.953
7-9am ZT22-0	18	0.01	1	1.18	0.278
9-11am ZT0-2	18	0.02	1	1.90	0.168
11-1pm ZT2-4	18	0.00	1	0.01	0.912
1-3pm ZT4-6	18	0.04	1	4.50	0.034*
3-5pm ZT6-8	29	0.01	1	0.65	0.422
5-7pm ZT8-10	29	0.01	1	0.87	0.351
7-9pm ZT10-12	29	0.01	1	0.93	0.335
9-11pm ZT12-14	29	0.02	1	2.00	0.157
11-1am ZT14-16	29	0.01	1	0.53	0.468
1-3am ZT16-18	29	0.00	1	0.01	0.911
3-5am ZT18-20	29	0.00	1	0.10	0.751
5-7am ZT20-22	29	0.01	1	0.82	0.365
7-9am ZT22-0	29	0.00	1	0.38	0.539
9-11am ZT0-2	29	0.01	1	1.19	0.725
11-1pm ZT2-4	29	0.03	1	3.39	0.066
1-3pm ZT4-6	29	0.01	1	1.08	0.299

Appendix 6.11 Pearson product moment correlations for transformed proportions of activity of Thr-Gly variants at 18°C and 29°C in various 2h windows(*, p<0.05)

CONSELVE		18°C			29°C		
Window		*r*	p	df	*r*	p	df
3-5pm	ZT6-8	0.01	0.904	137	0.17	0.067	119
5-7pm	ZT8-10	-0.12	0.179	137	-0.06	0.548	119
7-9pm	ZT10-12	-0.16	0.069	137	0.10	0.284	119
9-11pm	ZT12-14	-0.05	0.582	137	-0.24	0.010*	119
11-1am	ZT14-16	0.12	0.154	137	-0.17	0.068	119
1-3am	ZT16-18	0.13	0.120	137	-0.23	0.013*	119
3-5am	ZT18-20	0.13	0.134	137	-0.22	0.016*	119
5-7am	ZT20-22	0.11	0.211	137	-0.14	0.135	119
7-9am	ZT22-0	0.18	0.034*	137	-0.10	0.281	119
9-11am	ZT0-2	0.04	0.665	137	0.25	0.006*	119
11-1pm	ZT2-4	0.20	0.017*	137	0.32	0.000*	119
1-3pm	ZT4-6	0.15	0.090	137	0.15	0.114	119

Appendix 6.12 Pearson product moment correlations for transformed proportions of activity of Thr-Gly variants at 18°C and 29°C in various 2h windows(*, p<0.05)

LECCE		18°C			29°C		
Window		*r*	p	df	*r*	p	df
3-5pm	ZT6-8	0.22	0.021*	110	0.11	0.294	91
5-7pm	ZT8-10	0.08	0.407	110	0.12	0.255	91
7-9pm	ZT10-12	-0.04	0.719	110	-0.11	0.293	91
9-11pm	ZT12-14	0.01	0.925	110	-0.20	0.054	91
11-1am	ZT14-16	0.09	0.349	110	0.17	0.112	91
1-3am	ZT16-18	0.04	0.684	110	0.10	0.355	91
3-5am	ZT18-20	-0.01	0.938	110	0.13	0.216	91
5-7am	ZT20-22	0.00	0.988	110	0.04	0.745	91
7-9am	ZT22-0	-0.13	0.167	110	-0.01	0.961	91
9-11am	ZT0-2	-0.15	0.122	110	-0.01	0.911	91
11-1pm	ZT2-4	-0.00	0.983	110	0.21	0.051	91
1-3pm	ZT4-6	0.28	0.003*	110	0.07	0.485	91

APPENDICES FOR CHAPTER 7

Appendix.7.1. Autocorrelation, Planned comparisons between pooled Thr-Gly length alleles at different temperatures. (*, p<0.05; MS error, 0.402; df, 2299)

TEMP /°C	Thr-Gly LENGTH	MEAN PERIOD/h	planned comparison	difference between the means	MS	df	F	p
18	14	23.9	18-25	0.25	0.93	1	2.30	0.130
25	14	24.15	25-29	0.27	0.87	1	2.18	0.140
29	14	23.88						
18	17	23.95	18-25	0.24	4.99	1	12.44	0.000*
25	17	24.19	25-29	0.14	2.02	1	5.03	0.025*
29	17	24.05						
18	18	23.44	18-25	0.28	1.79	1	4.46	0.035*
25	18	23.72	25-29	0.24	1.64	1	4.08	0.043*
29	18	23.96						
18	20	23.78	18-25	0.39	18.41	1	45.82	0.000*
25	20	24.17	25-29	0.13	2.53	1	6.3	0.012*
29	20	24.04						
18	21	23.52	18-25	0.53	6.73	1	16.75	0.000*
25	21	24.05	25-29	0.03	0.02	1	0.06	0.810
29	21	24.08						
18	22	23.40	18-25	0.37	6.71	1	16.69	0.000*
25	22	24.03	25-29	0.04	0.03	1	0.08	0.773
29	22	24.07						
18	23	24.04	18-25	0.14	0.99	1	2.48	0.115
25	23	24.16	25-29	0.28	5.21	1	12.98	0.000*
29	23	23.88						

Appendix 7.2 Two way fixed ANOVA, Spectral (*, p<0.05)

Source	MS	df	F	p
Genotype(G)	6.41	6	15.43	0.000*
Temperature(T)	13.1	2	31.54	0.000*
T*G	2.3	12	5.53	0.000*
Error	0.415	2299		

Appendix 7.3. Spectral, Planned comparisons between pooled Thr-Gly length alleles at different temperatures. (*, p<0.05; MS error, 0.415; df, 2299)

TEMP /°C	Thr-Gly LENGTH	MEAN PERIOD/h	planned comparison	differences between the means	MS	df	F	p
18	14	24.21	18-25	0.05	0.03	1	0.080	0.78
25	14	24.26	25-29	0.02	0.01	1	0.013	0.91
29	14	24.24						
18	17	24.04	18-25	0.21	4.14	1	9.97	0.002*
25	17	24.25	25-29	0.13	1.75	1	4.22	0.040*
29	17	24.12						
18	18	23.40	18-25	0.38	3.42	1	8.24	0.004*
25	18	23.78	25-29	0.10	0.26	1	0.63	0.430
29	18	23.88						
18	20	23.87	18-25	0.43	22.31	1	53.72	0.000*
25	20	24.30	25-29	0.18	4.32	1	10.40	0.001*
29	20	24.12						
18	21	23.47	18-25	0.69	11.72	1	28.2	0.000*
25	21	24.16	25-29	0.14	0.58	1	1.39	0.240
29	21	24.02						
18	22	23.38	18-25	0.62	6.60	1	15.89	0.000*
25	22	24.00	25-29	0.18	0.83	1	2.00	0.160
29	22	24.18						
18	23	24.13	18-25	0.02	0.03	1	0.08	0.780
25	23	24.15	25-29	0.21	2.91	1	7.01	0.008*
29	23	23.94						

Appendix 7.4 Two way fixed ANOVA, Spectral (*, p<0.05)

Source	MS	df	F	p
Temperature(T)	22.34	2	61.5	0.000*
Line(L)	3.33	30	9.17	0.000*
T*L	1.61	60	4.43	0.000*
Error	0.363	2227		

Appendix 7.5 Correlations of Mean Locomotor Activity Period Difference for the Thr-Gly Length Alleles for Australia, Autocorrelation.

Temperature	Pearsons product moment	Spearmans rank order
29°C - 18°C	r=0.26, p=0.57	r_s=0.21, p=0.64
29°C - 25°C	r=0.13, p=0.78	r_s=0.04, p=0.94
25°C - 18°C	r=0.33, p=0.47	r_s=0.21, p=0.64

Appendix 7.6 Correlations of mean locomotor activity period difference for the Thr-Gly length alleles for Australia, Spectral.

Temperature	Pearsons product moment	Spearmans rank order
29°C - 18°C	r=0.25, p=0.59	r_s=0.21, p=0.64
29°C - 25°C	r=0.14, p=0.77	r_s=-0.3, p=0.48
25°C - 18°C	r=0.41, p=0.37	r_s=0.21, p=0.64

Appendix 7.7 Autocorrelation, Pearsons product moment correlations for period length of Thr-Gly variants at 18°C, 25°C and 29°C(*, p<0.05)

	18°C			25°C			29°C		
Thr-Gly Variant	r	p	df	r	p	df	r	p	df
$(Thr\text{-}Gly)_{17}$	-0.09	0.272	159	0.11	0.127	195	-0.14	0.045*	210
$(Thr\text{-}Gly)_{20}$	-0.16	0.015*	232	-0.00	0.948	257	-0.04	0.480	330

Autocorrelation, Spearman rank order correlations for period length of Thr-Gly variants at 18°C, 25°C and 29°C(*, p<0.05)

	18°C			25°C			29°C		
Thr-Gly Variant	r_s	p	df	r_s	p	df	r r_s	p	df
$(Thr\text{-}Gly)_{17}$	-0.09	0.277	159	0.10	0.162	195	-0.02	0.793	210
$(Thr\text{-}Gly)_{20}$	-0.21	0.001*	232	-0.05	0.432	257	-0.02	0.686	330

Appendix 7.8 Spectral, Pearsons product moment correlations for period length of Thr-Gly variants at 18°C, 25°C and 29°C(*, p<0.05)

	18°C			25°C			29°C		
Thr-Gly Variant	*r*	p	df	*r*	p	df	*r*	p	df
$(Thr\text{-}Gly)_{17}$	-0.01	0.931	159	0.15	0.031*	195	-0.14	0.043*	210
$(Thr\text{-}Gly)_{20}$	-0.14	0.038*	232	0.02	0.699	257	-0.15	0.006*	330

Spectral, Spearman rank order correlations for period length of Thr-Gly variants at 18°C, 25°C and 29°C(*, p<0.05)

	18°C			25°C			29°C		
Thr-Gly Variant	r_s	p	df	r_s	p	df	r_s	p	df
$(Thr\text{-}Gly)_{17}$	0.04	0.613	159	0.16	0.027*	195	-0.05	0.466	210
$(Thr\text{-}Gly)_{20}$	-0.20	0.002*	232	-0.04	0.506	257	-0.13	0.021*	330

REFERENCES

Akashi, H. (1994) Synonymous codon usage in *D. melanogaster* - Natural selection and translational accuracy. Genetics 136 : 927-935.

Akashi, H. (1995) Inferring weak selection from patterns of polymorphism and divergence at silent sites in *Drosophila* DNA. Genetics 139 : 1067-1076.

Aquadro, C.F. (1992) Why is the genome variable ? Insights from *Drosophila.* Trends in Genetics 8 : 355-362.

Anderson, P.R. (1981) in Genetic Studies of *Drosophila* populations. Edited by J.B. Gibson and J.G. Oakeshott. Australian National University Press, Canberra: pp237-250.

Anderson, P.R. and Oakeshott J.G. (1984) Parallel geographic patterns of allozyme variation in two sibling *Drosophila* species. Nature 308 : 729-731.

Anderson, S. and Lambertsson, A. (1993) Evolution of the *dec-1* eggshell locus in *Drosophila.* II. Intraspecific DNA sequence analysis reveals length mutations in a repetitive region in *D. melanogaster.* J. Mol. Evol. 36 : 536-544.

Aronson B.D., Johnson K.A., Loros J.J. and Dunlap J.C. (1994) Negative feedback defining a circadian clock: autoregulation of the clock gene *frequency.* Science 263 : 1578-1584.

Arpaia, G., Loros, J.J., Dunlap, J.C., Morelli, G. and Macino, G. (1993) The interplay of light and the circadian clock- independent dual regulation of clock controlled gene *ccg-2* (*eas*). Plant Physiology 102 : 1299-1305.

Asghari, V., Schoots, O., Vankats, S., Ohara, K., Jovanovic, V., Guan, H.C., Bunzow, J.R., Petronis, A. and Van Tol, H.H.M. (1994) Dopamine D4 receptor repeat - analysis of different native and mutant forms of human and rat genes. Journal of Molecular Pharmacology 46 : 364-373.

Ayala, F.J., Serra, L. and Prevosti, A. (1989). A grand experiment in evolution: the *Drosophila subobscura* colonization of the Americas. Genome 31 : 246-255.

Bachmann, B., Luke, W., Hunsmann, G. (1990). Improvement of PCR amplified DNA sequencing with the aid of detergents. Nucleic Acids Res. 18 : 1309.

Bargiello T.A. and Young M.W. (1984a) Molecular genetics of a biological clock in *Drosophila.* Proc. Natl. Acad. Sci. USA. 81 : 2142-2146.

Bargiello T.A., Jackson F.R. and Young M.W. (1984b) Restoration of circadian behavioural rhythms by gene transfer in *Drosophila.* Nature 312 : 752-754.

Bargiello T.A., Saez L., Bayles M.K., Gasik G., Young M.W. and Spray D.C. (1987) The *Drosophila* clock gene *per* affects intercellular junctional communication. Nature 328 : 686-691.

Baylies M.K., Bargiello T.A., Jackson F.R and Joung M.W. (1987) Changes in abundance or structure of the *per* gene product can alter periodicity of the *Drosophila* clock. Nature 326 : 390-392.

Baylies M.K., Vosshall L.B., Sehgal A. and Young M.W. (1992). New short period mutations of the *Drosophila* clock gene *per*. Neuron 9 : 575-581.

Begun, D.J. and Aquadro, C.F. (1991). Molecular population genetics of the distal portion of the *X* chromosome in *Drosophila*: Evidence for genetic hitchhiking of the *yellow-achaete* region. Genetics 129 : 1147-1158.

Begun, D.J. and Aquadro, C.F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356 : 519-520.

Begun, D.J. and Aquadro, C.F. (1993). African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature 365 : 548-550.

Benjamin, J., Li, L., Patterson, C., Greenberg, B.D., Murphy, D.L. and Hamer, D.H. (1996) Nature Genetics 12 : 81-84

Ben-Shlomo R., Ritte U. and Nevo E. (1996). Circadian rhythm and the *per* ACNGGN repeat in the mole rat, *Spalax ehrenbergi*. Behaviour genetics 26 : 177-184.

Berry, A.J., Aijoka, J.W. and Kreitman, M. (1991). Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. Genetics 129 : 1111-1117.

Bertholf L.M. (1932). The extend of the spectrum for *Drosophila* and the distribution of stimulative in it. Z. Vergl. Physiol. 18 : 32-64.

Brandes, C. , Plautz, J.D., Stanewsky, R., Jamison, C.F., Straume, M., Wood, K.V., Kay, S.A. and Hall, J.C. (1996) Novel features of *Drosophila period* transcription revealed by realtime luciferase reporting. Neuron 16 : 687-692.

Brasseil K.E and Reif D. (1979) A procedure to generate Theissen polygons. Geographical Analysis 11 : 289-303.

Bruce, V.G. and Pittendrigh, C.S. (1957) Endogenous rhythm in insects and microorgansims. American Naturalist 91 : 179-195.

Bruce, V.G. (1970) The biological clock of *C reinhardi*. J Protozool. $(Thr\text{-}Gly)_{17}$: 328-333.

Bruce, V.G. (1972) Mutants of the biological clock in *C reinhardi*. Genetics 70 : 537-548.

Burbach K.M., Poland, A. and Bradfield C.A. (1992) Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. Proc. Natl. Acad. Sci. USA 89 : 8185-8189.

Cann, R.L., Stoneking, M. and Wilson, A.C. (1987) Mitochondrial DNA and human evolution. Nature 325 : 31-36.

Castiglione-Morelli M.A., Guantieri V., Villani V., Kyriacou C.P., Costa R. and Tamburro A.M. (1995) Conformational study of the Thr-Gly repeat in the *Drosophila* clock protein PERIOD. Proc. R. Soc. London 260 : 155-163

Charlesworth, B. (1990) The Evolutionary genetics of adaptation. In Evolutionary Innovations edited by M.H. Nitecki, University of Chicago Press: pp47-70.

Charlesworth B., Sneigowski P. and Stephen W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 257 : 588-590.

Chatfield C. (1980) The analysis of time series. Chapman and Hall, London

Citri Y., Colot H.V., Jacquier A.C., Yu Q., Hall J.C., Baltimore, D. and Rosbash M. (1987) A family of unusually spliced biologically active transcripts encoded by a *Drosophila* clock gene. Nature 326 : 42-47.

Colot, H.V., Hall, J.H. and Rosbash M. (1988) Interspecific comparison of the *period* gene of *Drosophila* reveals large blocks of non-conserved coding DNA. EMBO J. 7 : 3929-3937.

Costa R.A., Peixoto, A.A., Barbujani, G. and Kyriacou, C.P. (1992) A latitudinal cline in a *Drosophila* clock gene. Proc. R. Soc. London 250 : 43-49.

Costa R.A., Peixoto, A.A., Thackeray, J.R., Dalgleish, R. and Kyriacou, C.P. (1991) Length polymorphism in the threonine-glycine encoding repeat region of the *period* gene in *Drosophila*. J. Mol. Evol. 32 : 238-246.

Coté, G.G. and Brody, S. (1986) Circadian rhythms in *Drosophila melanogaster*: analysis of period as a function of gene dosage at the *per* (*period*) locus. J. Theor. Biol. 121 : 487-503.

Cowling, D.E. and Burnet, B. (1981) Courtship songs and genetic control of their acoustic characteristics in sibling species of the *D. melanogaster* subgroup. Anim. Behav. 29 : 924-935.

Cowling, R.M., Rundel, P.W., Lamont, B.B., Arroyo, M.K. and Arianoutsou, M. (1996) Plant diversity in Mediterranean-climate regeions. Trends in Ecology and Evolution 11 : 362-366.

Coyne, J.A. (1992) Gentics and Speciation. Nature 355 : 511-515.

Crews S.T., Thomas J.B. and Goodman C.S. (1988) The *Drosophila single-minded* gene encodes a nuclear protein with sequence similarity to the *per* gene product. Cell 52 : 143-151.

Crosthwaite, S.K., Loros, J.J. and Dunlap, J.C. (1995) Light-inducesd resetting of a circadian clock is mediated by a rapid increase in *frequency* transcript. Cell 81 : 1003-1012.

Curtin, K.D., Huang, Z.J. and Rosbash, M. (1995) Temporally regulated nuclear entry of the *Drosophila period* protein contributes to the circadian clock. Neuron 14 : 365-372.

David, J.R. and Capy, P. (1988) Genetic variation of *Drosophila melanogaster* natural populations. Trends in Genetics 4 : 106-111.

David, J.R., Alonso-Moraga, A., Borai, F., Capy, P., Mercot, H., McEvey, S.F., Munoz-Serrano, A. and Tsakas S. (1989) Latitudinal variation of *Adh* gene frequencies in *Drosophila melanogaster*: a Mediterranean instability. Heredity 62 : 11-16.

Demets, R, Tomson, A.M., Stegwee, D. and Van den Ende, H. (1987) Control of the mating competence rhythm in *C eugametos*. J. Gen. Microbiol. 133 : 1081-1088.

Diggle, P.J. (1990) Time Series: A biostatistical introduction. Oxford Science Publications. Clarendon Press, Oxford.

Djian, P. and Green, H. (1991) Involucrin gene of tarsioids and other primates: Alternatives in evolution of the segment of repeats. Proc. Natl. Acad. Sci. USA. 88 : 5321-5325.

Djian, P., Phillips, M., Easley, K., Huang, E., Simon H., Rice R.H. and Green H. (1993) The involucrin genes of the mouse and the rat; study of their shared repeats. Molecular Biology and Evolution 10 : 1136-1149.

Dobzhansky, T . (1937) Genetics and the origin of species. Columbia University Press, New York.

Dover, G.A. (1987) DNA turnover and the Molecular Clock. J. Mol. Evol. 26 : 47-58.

Dover, G.A. (1989) Slips, Strings and Species. Trends in Genetics 5 : 100-103.

Dowse, H.B. and Ringo, J.M. (1989) The search for hidden periodicities in biological time series revisited. J. Theor. Biol. 139 : 487-515.

Dowse, H.B., Hall, J.C. and Ringo, J.M. (1987) Circadian and ultradian rhythms in *period* mutants of *Drosophila melanogaster*. Behaviour. Genetics. 17 : 19-35.

Dunlap J.C. (1993). Genetic analysis of circadian clocks. Annu. Rev. Physiol. 55 : 683-728.

Dushay, M.S., Konopka, R. J., Orr, D., Greenacre, M., Kyriacou, C.P., Rosbash, M. and Hall, J.C. (1990) Phenotypic and genetic analysis of *Clock*, a new circadian rhythm mutant in *Drosophila melanogaster*. Genetics 125 : 557-558.

Dushay, M.S., Rosbash, M. and Hall, J.C. (1989) The *disconnected* visual system mutations in *Drosophila melanogaster* drastically disrupt circadian rhythms. J. Biol. Rhythms 4 : 1-27.

Easteal, S . (1985) The ecological genetics of introduced populations of the giant toad, *Bufo marinus*. III Geographical patterns of variation. Evolution 39 : 1065-1075.

Ebstein, R.P., Novick, O ., Umansky, R., Priel, B., Osher,Y., Blaine, D., Bennett, E.R., Nemanov, L., Katz, M. and Belmaker, R.H. (1996) Dopamine D4 receptor (*D4DR*) exon III polymorphism asscociated with the human personality trait of novelty seeking. Nature Genetics 12 : 78-80

Edery, I., Rutila, J.E. and Rosbash, M. (1994a) Phase-shifting of the circadian clock by induction of the *Drosophila period* protein. Science 263 : 237-240.

Edery, I., Zwiebel, L.J., Dembinska, M.E. and Rosbash, M. (1994b) Temporal phosphorylation of the *Drosophila* period protein. Proc. Natl. Acad. Sci. USA 91 : 2260-2264.

Eichler, E.E., Kunst, C.B, Ligenbeel, K.A, Ryder, O.A, Davison , D., Warren, S.T., Nelson, D.L. (1995) Evolution of the cryptic *FMR1* CGG repeat. Nature Genetics 11 : 301-308

Eickbush, T.H. and Burke, W.D. (1986) The silkmoth late chorion locus II. Gradients of gene conversion in two paired multigene families. J. Mol. Biol. 190 : 357-366.

Endler, J.A (1977) Geographic variation, speciation and clines. Princeton University Press.

Endler, J.A. (1986) Natural selection in the wild. Princeton University Press.

Epenshade, E.B. and Morrison, J.L. (1986) Goodes World Atlas. 17th Edition. Random and McNally. New York.

Ewer, J., Frisch, B., Hamblen-Coyle, M.J., Rosbash, M. and Hall, J.C. (1992) Expression of the *period* clock gene within different cell types in the brain of *Drosophila* adults and mosaic analysis of these cells' influence on circadian behavioral rhythms. J. Neurosci. 12 : 3321-3349.

Ewer, J., Hamblen-Coyle, M., Rosbash, M. and Hall, J.C. (1990) Requirement for *period* gene expression in the adult and not during development for locomotor activity rhythms of imaginal *Drosophila melanogaster*. J. Neurogenet. 7 : 31-73.

Excoffier, L., Smouse, P.E., Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 13 : 479-491.

Feldman, J.F. and Hoyle, M.N. (1973). Isolation of circadian clock mutants of *Neurospora crassa*. Genetics 75 : 605-613.

Fisher, R.A. (1930) The Genetical theory of natural selection. Clarendon Press, Oxford

Fleissner, G., Fleissner, G. and Frisch, B. (1993) A new-type of putative nonvisual photoreceptors in the optic lobe of beetles. Cell And Tis. Res. 273 : 435-445.

Flint, K., Rosbash, M. and Hall, J.C. (1993). Transfer of dye among salivary gland cells is not affected by genetic variations of the *period* clock gene in *Drosophila melanogaster*. J. Membrane Biol. 136 : 333-342.

Ford, M.J., Yoon, C.K. and Aquadro, C.F. (1994) Molecular evolution of the *period* gene in *Drosophila athabasca*. Molecular Biology And Evolution 11 : 169-182.

Frisch, B., Hardin, P.E., Hamblen-Coyle, M.J., Rosbash, M. and Hall, J.C. (1994) A promoterless *period* gene mediates behavioral rhythmicity and cyclical *per* expression in a restricted subset of the *Drosophila* nervous system. Neuron 12 : 555-570.

Garcia-Fernandez, J.M., Siwicki, K.K and Foster, R.G. (1994) *per*- like immunoreactive neurons within the brain of *Xenopus* innervate both the retinal and pineal organ . In Advances in Pineal Research, edited by M. Moller and P Pevet: pp19-24.

Gekakis, N., Saez, L., Delahayebrown, A.M., Myers, M.P., Sehgal, A., Young, M.W. and Weitz, C.J. (1995) Isolation of *timeless* by *per* protein interaction, defective interaction between *timeless* protein and long-period mutant per^{L} . Science, 270 : 811-815.

Gentilli, J. (1971) World survey in Climatology: Climates of Australia and New Zealand, 13. Elsevier Publishing company, Amsterdam.

Gibson, G and Hogness, D.S. (1996) Effect of polymorphism in the *Drosophila* regulatory gene *Ultrabithorax* on homeostatic stability. Science 271 : 200-203.

Ginsburg, P.E. (1993) The Shadow of Death- The hunt for a serial killer, The Berkley Publishing Group., USA.

Gloor, G. and Engels, W. (1990) Single-fly DNA preps for PCR. Drosophila Information Newsletter 1.

Gray, I. and Jeffreys, A.J. (1991) Evolution transience of hypervariable minisatellites in man and primates. Proc R. Soc. London. Biol. 243 : 241-253.

Green, H. and Djian, P. (1992) Consecutive actions of different gene-altering mechanisms in the evolution of involucrin. Molecular Biology Evolution 9 : 977-1017.

Greenacre, M.L., Ritchie, M.G., Byrne, B.C. and Kyriacou, C.P. (1993) Female Song Preference and the period Gene In *Drosophila.* Behaviour Genetics $(Thr\text{-}Gly)_{23}$: 85-90.

Griffith, L.C., Verselis, L.M., Aitken, K.M., Kyriacou, C.P., Danho, W. and Greenspan, R.J. (1993) Inhibition of calcium calmodulin dependent protein kinase in *Drosophila* disrupts behavioural plasticity. Neuron 10 : 501-509.

Hale, R. (1994) The role of humiliation and embarrassment in serial murder. Journal of Psychology 31 : 17-23

Hall, J.C. (1990). Genetics of circadian rhythms. Ann. Rev. Genet. 24 : 659-697.

Hall, J.C. and Kyriacou, C.P. (1990) Genetics of biological rhythms in *Drosophila.* Adv Ins Physiol 22 : 221-298.

Hamblen, M., Zehring, W.A., Kyriacou, C.P., Reddy, P., Yu, Q., Wheeler, D.A., Zwiebel, L.J., Konopka, R.J., Rosbash, M. and Hall, J.C. (1986) Germ-line transformation involving DNA from the *period* locus in *Drosophila melanogaster*: overlapping genomic fragments that restore circadian and ultradian rhythmicity to per^0 and per^- mutants. J. Neurogenet. 3 : 249-291.

Hamblen-Coyle, M., Konopka, R.J., Zweibel, L.J., Colot, H.V., Dowse, H.B., Rosbash, M. and Hall, J.C. (1989) A new mutation at the *period* locus of *Drosophila melanogaster* with some novel effects on circadian rhythms. J. Neurogenet. 5 : 229-256.

Hamblen-Coyle, Wheeler, D.A., Rutila, J.E. Rosbash, M. and Hall, J.C. (1992) Behaviour of period-altered circadian rhythms mutants of Drosophila in light:dark cycles (Diptera: Drosophilidae). J. Insect Behav. 5 : 417-446.

Hardin, P.E., Hall, J.C. and Rosbash, M. (1990) Feedback of the *Drosophila period* gene product on circadian cycling of its messenger RNA levels. Nature 343 : 536-540.

Hardin, P.E., Hall, J.C. and Rosbash, M. (1992) Circadian oscillations in *period* gene mRNA levels are transcriptionally regulated. Proc. Natl. Acad. Sci. USA 89 : 11711-11715.

Hardin, P. and Siwicki, K.K. (1995) The Multiple Roles Of *per* In the *Drosophila* Circadian Clock. Seminars In the Neurosciences 7 : 15-25.

Harding, R.M., Boyce, A.J. and Clegg, J.B. (1992) The evolution of tandemly repetitive DNA: recombination rules. Genetics 132 : 847-859.

Helfrich-Forster, C and Homberg, U. (1993) Pigment dispersing hormone immunoreactive neurons in the nervous system of *D. melanogaster* and of several mutants with altered circadian rhythmicity. (1993) J. Comp. Neurol. 337 : 177-190.

Hennig W. (1981) Insect phylogeny. Pitman Press, Bath, England

Hill, W.G.(1982) Rates of change in quantitative traits from fixation of new mutations. Proc. Natl. Acad. Sci. USA 79 : 142-145.

Hoffman, E.C., Reyes, H., Chu, F.F., Sander, F., Conley, L.H., Brooks, B.A. and Hankinson, O. (1991). Cloning of a factor required for activity of the Ah (dioxin) receptor. Science 252 : 954-958.

Housemann, D. (1995) Gain of glutamines gain of function. Nature Genetics 10 : 3-4.

Huang, Z.J., Curtin, K.D. and Rosbash, M. (1995) PER protein interactions and temperature compensation of a circadian clock in *Drosophila*. Science 267 : 1169-1172.

Huang, Z.J., Edery, I. and Rosbash, M. (1993) PAS is a dimerization domain common to *Drosophila* Period and several transcription factors. Nature 364 : 259-262.

Hudson, R.R., Kreitman,M. and Aguade, M. (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116 : 153-159.

Hulme, M., Conway, D., Jones, P..D., Jiang T., Barrow, E.M. and Turney, C. (1995) Construction of a 1961-90 European climatology for climate change modelling and impact applications. International Journal of Climatology 15 : 1333-1363.

Hunter-Ensor, M., Ousley, A. and Sehgal, A. (1996) Regulation of the *Drosophila* protein *timeless* suggests a mechanism for resetting the circadian clock by light. Cell 84 : 677-685.

Ishida, N., Mitsui, Y. and Tsukada, M. (1994) Molecular conformation and physical properties of poly(Gly-Thr)$_n$: a compound model for the *period* repetitive sequences. Mol. Biol. Life Sci. Adv. 13 : 107-111.

Iwasaki, K. Liu, D.W.C. and Thomas, J.C. (1995) Genes that control a temperature compensated ultradian clock in *C elegans*. Proc. Natl. Acad. Sci. USA 92 : 10317-10321.

Jackson, F.R., Bargiello, T.A., Yun, S.H., Young, M.W. (1986) Product of *per* locus of *Drosophila* shares homology with proteoglycans. Nature 320 : 185-188.

Jackson, F.R. (1993). Circadian rhythm mutants of *Drosophila*. In Molecular Genetics of Biological Rhythms, edited by M.W. Young. Marcel Dekker, New York: pp. 91-121.

James, A. and Partridge, L. (1995) Thermal evolution of rate of larval development in *D. melanogaster* in laboratory and field populations. J. Evol. Biol. 8 : 315-330.

Jeffreys, A.J., Royle, N.J., Wilson, V. and Wong, Z. (1988a) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature 332 : 278-281

Jeffreys, A.J., Wilson, V., Neumann, R. and Keyte, J. (1988b) Amplification of human minisatellites by polymerase chain reaction: towards DNA fingerprinting of single cells. Nucleic Acid Res. 16 : 10953-10971.

Jeffreys, A.J., Monckton, D.G., Tamaki, K., Neil, D.L., Armour, J.A.L., MacLeod, A., Collick, A., Allen, M. and Jobling, M. (1993) Minisatellite variant mapping: Application to DNA typing and mutation analysis. In: DNA Fingerprinting: State of the Science edited by S.D.J. Pena, R. Chakraborty, J.T. Epplen and A.J. Jeffreys. Birkhauser Verlag, Basel: pp. 125-139.

Jeffreys, A.J., Tamaki, K., Macleod, A., Monckton, G., Neil, D.L. and Amour, J.A.L. (1994) Nature Genetics 6 : 136-145.

Jeffreys, A.J., Neumann, R., and Wilson, V. (1990) Repeat unit sequence variation in minisatellites: A novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell 60:473-485.

Kalmus, H. (1935) Periodizitat und Autichionie (Ideochronie) als Zeitregelnde Eigenschaffen der Organismen. Biol. Generalis 11 : 93-114.

Kimura, M. (1983). The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Kliman, R.M. and Hey, J. (1993) DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. Genetics 133 : 375-387.

Kondo, T., Strayer, C.A., Kulkarni, R.D., Taylor, W., Ishiura, M., Golden, S.S. and Johnson, C.H. (1993) Circadian rhythms in prokaryotes: Luciferase as a reporter of circadian gene expression in cyanobacteria. Proc. Natl. Acad. Sci. USA. 90 : 5672-5676.

Kondo, TIM ., Tsinoremas, N.F., Golden, S.S., Johnson, C.H. Kutsuna, S . and Ishiura, M. (1994) Circadian clock mutants of cyanobacteria. Science 266: 1233-1236.

Konopka, R.J. (1972) Ph.D. thesis, California Institute of Technology, Pasadena, California, USA.

Konopka, R.J. and Benzer, S. (1971) Clock mutants of *Drosophila melanogaster*. Life Sci. Adv. Series C 5 : 47-49.

Konopka, R.J., Pittendrigh, C. and Orr, D. (1989) Reciprocal behaviour associated with altered homeostasis and photosensitivity of *Drosophila* clock mutants. J. Neurogenet. 6 : 1-10.

Konopka, R.J., Smith, R. and Orr, D. (1991) Characterisation of *Andante*, a new Drosophila clock mutant, and its interaction with other clock mutants. J. Neurogenet. 7 : 103-114.

Konopka, R.J., Hamblen-Coyle, M.J., Jamison, C.F. and Hall, J. C. (1995) An ultrashort clock mutation at the *period* locus of *D. melanogaster* that reveals some new features of the flies circadian system. Journal of Biological Rhythms 9 : 189-216.

Kornhauser, J.M., Mayo, J.M. and Takahashi, J.S. (1996) Light, immediate early genes and circadian rhythms. Behaviour Genetics 26 : 221-240.

Kreitman M. (1991) Detecting selection at the level of DNA. In: Evolution at the molecular level edited by R.K. Selander, A.G. Clark and T.S. Whittam. Sinauer Ass. Inc. USA.

Kreitman, M. (1996) The neutral theory is dead-long live the neutral theory. BioEssays, 18, 678-683.

Kuhl, D.P.A. and Caskey, C.T. (1993) Trinucleotide repeats and genome variation. Current Opinion in Genetics & Development 3 : 404-407.

Kyriacou, C.P. and Hall, J.C. (1980) Circadian rhythms mutations in *Drosophila melanogaster* affect short term fluctuations in the male's courtship song. Proc. Natl. Acad. Sci. USA 77 : 6929-6933.

Kyriacou, C.P. and Hall, J.C. (1982) The function of courtship song rhythms in *Drosophila.* Anim. Behav. 30 : 794-801.

Kyriacou, C.P. and Hall, J.C. (1986) Interspecific genetic control of courtship song production and reception in *Drosophila.* Science 232 : 494-497.

Kyriacou, C.P. and Hall, J.C. (1989) Spectral analysis of *Drosophila* courtship song rhythms. Anim. Behav. 37 : 850-859.

Kyriacou, C.P., Oldroyd, M., Wood, J., Sharp, M. and Hill, M.. (1990a) Clock mutations alter developmental timing in *Drosophila.* Heredity 64 : 395-401.

Kyriacou, C.P., Vandenberg, M.J. and Hall, J.C. (1990b) Drosophila courtship song cycles in normal and *period* mutant males revisited Behaviour Genetics 20 : 617-644

Kyriacou, C.P. (1990) The molecular ethology of the *period* gene in *Drosophila.* Behaviour Genetics 20 : 191-211.

Kyriacou, C.P. and Burnet, B. (1979) Genetic analysis of the phototaxis near the upper limit of the visual spectrum of *Drosophila melanogaster.* Behaviour Genetics 9 : 123-128.

Kyriacou, C.P., Oldroyd, M., Wood, J., Sharp, M. and Hill, M. (1990a) Clock mutations alter developmental timing in *Drosophila.* Heredity 64 : 395-401.

Kyriacou, C.P., van den Berg, M.J. and Hall, J.C. (1990b) *Drosophila* courtship song cycles in normal and *period* mutant males revisited. Behaviour Genetics 20 : 631-658.

Lande, R. (1983) The response to selection on major and minor mutations affecting a metrical trait. Heredity 50 : 47-65.

Lee, C., Parikh, T., Itsukaichi, T., Bae, K. and Edery, I. (1996) Resetting the *Drosophila* clock by photic regulation of PER and PER-TIM complex. Scinece 271 : 1740-1744.

Levine, J.D., Sauman, I., Imbalzano, M., Reppert, S.M. and Jackson, F.R. (1995) Period protein from the giant silkmoth *Antheraea pernyi* functions as a circadian clock element in *Drosophila melanogaster.* Neuron 15 : 147-157.

Levinson, G. and Gutman, G.A. (1987) Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Molecular Biology Evolution 4 : 203-221.

Li-Weber, M, Groot, E.J., Schweiger, H.G. (1987) Sequence homology to the *Drosophila per* locus in higher plant nuclear DNA and in *Acetabularia* chloroplast DNA. Molecular General Genetics 209 : 1-7.

Liu X., Lorenz L., Yu Q., Hall J.C. and Rosbash M (1988) Spatial and temporal expression of the *period* gene in *Drosophila melanogaster*. Genes Dev. 2 : 228-238.

Liu, X. and Thomas, J.H. (1994) Regulation of a periodic motor program in *C elegans*. J. Neurosci. 14 : 1953-1962.

Liu, X., Zwiebel, L.J., Hinton, D., Benzer, S., Hall, J.C. and Rosbash, M. (1992) The *period* gene encodes a predominantly nuclear protein in the adult *Drosophila*. J. Neurosci. 12 : 2735-2744.

Liu, Y., Tsinoremas, N., Johnson, Clebedeva, N., Golden, S., Ishiura, M. and Kondo, T. (1995) Cicadain orchestration of gene expression in cyanobacteria. Genes. Dev. 9 1469-1478.

Loros, J.J., Denome, S.A., Dunlap, J.C. (1989) Molecular cloning of genes under the control of the circadian clock in *Neurospora*. Science 243 : 385-388.

Loros, J.J., Lichens-Park, A., Lindgren, K.M. and Dunlap, J.C. (1993) Molecular genetics of genes under circadian temporal control in *Neurospora*. In Molecular Genetics of Biological Rhythms edited by M.W. Young. Marcel Dekker, New York: pp. 55-72.

Loros, J.J. (1995) The molecular basis of the Neurospora clock. Seminars in the Neurosciences 7 : 3-13.

Matsui, M., Mitsui, Y. and Ishida, N. (1993) Circadian regulation of *per* repeat messenger RNA in the suprachiasmatic nucleus of the rat brain. Neurosci. Lett. 163 : 189-192.

Matsumoto, A., Motoshige, T ., Murata, T ., Tomioka, K., Tanimura, T . and Chiba, Y. (1994) Chronobiological analysis of a new clock mutant *Toki*, in *D. melanogaster* . J. Neurogenet. 9 : 141-155.

Maynard-Smith, J. and Haigh, J. (1974) The hitch-hiking effect of a favourable gene. Genet. Res. 23 : 23-35.

McClung, C.R., Fox, B.A. and Dunlap, J.C. (1989) *frequency*, a clock gene in *Neurospora* shares a sequence element with the *Drosophila* clock gene *period*. Nature 339 : 558-562.

McDonald, J.F. and Ayala, F.J. (1974) Genetic response to environmental heterogeneity. Nature 250 : 572-574.

McKenzie, J.A. and Batterham, P . (1994) The genetic, molecular and phenotypic consequences of selection for insecticide resistance. Trends in Ecology and Evolution 9 : 166-169.

Menozzi, P ., Piazza, A., Cavalli-Sforza, L.L. (1978) Synthetic maps of human gene frequencies in Europeans. Science 201 : 786-792.

Merrow, M.W. and Dunlap, J.C. (1994) Intergeneric complementation of a circadian rhythmicity defect - phylogenetic conservation of structure and function of the clock gene-frequency. Embo Journal 13 : 2257-2266.

Millar, A.J., Straume, M., Chory, J., Chau, N.H. and Kay, S. (1995a) Regulation of circadian period by phototransduction pathways in Arabidopsis. Science 267 : 1163-1166.

Millar, A.J., Carre, I.A., Strayer, C.A., Chua, N.H. and Kay, S. (1995b) Circadian clock mutants in Arabidopsis identified by luciferase imaging. Science 267 : 1161-1163.

Muller, H.J. (1940) Bearings of the *Drosophila* work on systematics. In The new systematics edited by J.S. Huxley, Clarendon Press, Oxford: pp185-268.

Murata, T., Matsumoto, A., Tomioka, K. and Chiba, Y. (1995) *ritsu* - a rhythm mutant from a natural-population of *Drosophila melanogaster*. J. Neurogenet. 9 : 239-249.

Myers, M.P., Wagner-Smith, K., Wesley, C.S., Young, M. and Sehgal, A. (1995) Positional cloning and sequence analysis of the *Drosophila* clock gene, *timeless*. Science 270 : 805-808.

Newby, L.M. and Jackson, F.R. (1991) *Drosophila ebony* mutants have altered circadian activity rhythms but normal eclosion rhythms. J. Neurogenet. 7 : 85-101.

Newby, L.M., White, L., DiBartolomeis, S.M., Walker, B.J., Dowse, H.B., Ringo, J.M., Khuda, N. and Jackson, F.R. (1991). Mutational analysis of the *Drosophila miniature-dusky* (m-dy) locus: effects of cell size and circadian rhythms . Genetics 128 : 571-582.

Nielsen, J., Peixoto, A.A., Piccin, A., Costa, R., Kyriacou, C.P. and Chalmers, D. (1994) Big Flies, Small Repeats - the Thr-Gly Region Of the Period Gene In Diptera. Molecular Biology and Evolution 11 : 839-853.

Oakeshott, J.G. (1979) Selection affecting enzyme polymorphism in laboratory populations of *D. melanogaster*. Oecologia 43 : 341-354.

Oakeshott, J.G., Chambers, G.K., Gibson, J.B., and Willcocks, D.A. (1981) Latitudinal relationships of Esterase-6 and phosphoglucomutase gene frequencies in *D. melanogaster*. Heredity 47 : 385-396.

Oakeshott, J.G., Gibson, J.B., Anderson, P.R., Knibb, W.R., Anderson, D and Chambers, G.K. (1982) Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. Evolution 36 : 88-96.

Oden N.L. (1984) Assessing the significance of a spatial correlogram. Geographical Analysis 16 : 1-16.

Ohta, T. (1996) The current significance and standing of the neutral and nearly neutral theories. BioEssays 18 : 673-677.

Palopoli, K. and Patel, N.H. (1996) Neo-Darwinian developmental evolution: can we bridge the gap between pattern and process? Current Opinions in Genetics and Development, 6, 502-508.

Peixoto, A.A (1993a) Molecular evolution of a repetitive region within a clock gene in Drosophila. PhD Thesis, University of Leicester.

Peixoto, A.A., Campesan, S., Costa, R. and Kyriacou, C.P. (1993b) Molecular evolution of a repetitive region within the *per* gene of *Drosophila*. Molecular Biology and Evolution 10 : 127-139.

Peixoto, A.A., Costa, R., Wheeler, J.C., Hall, J.C. and Kyriacou, C.P. (1992) Evolution of the threonine-glycine repeat region in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Evol. 35 : 711-719.

Petersen, G., Hall, J.C. and Rosbash, M. (1988) The *period* gene of *Drosophila* carries species-specific behavioural instructions. EMBO Journal. 7 : 3939-3947.

Piazza, A., Menozzi, P., Cavalli-Sforza, L.L. (1981) Synthetic gene frequency maps of man and selective effects of climate. Proc. Natl. Acad. Sci. USA 78 : 2638-2642.

Pittendrigh, C.S. (1960) Circadian rhythms and the circadian organisation of living systems. Cold Spring Harbour Symp. Quant. Biol. 25 : 159-184.

Pittendrigh, C.S., Bruce, V.G., Rosenzweig, N.S. and Rubin, M.L. (1959) A biological clock in *Neurospora*. Nature 184 : 169-170.

Pittendrigh, C.S. (1954) On the temperature independence in the clock system controlling emergence time in *Drosophila*. Proc. Natl. Acad. Sci. USA 40 : 1018-1029.

Pittendrigh, C.S. (1965) On the mechanism of entrainment of a circadian rhythm by light cycles. In Circadian clocks edited by J. Aschoff. North Holland, Amsterdam: pp277-297

Pittendrigh, C.S. (1974) Circadian oscillation in cells and the circadian organisation of multicellular systems. In: The Neurosciences, Third Study Program edited by F.O. Schmitt and F.R. Worden. MIT Press, Cambridge, Mass: pp. 437-458.

Pittendrigh, C.S. (1981) Circadian oscillations in cells and the circadian organisation of multicellular systems. In Biological clocks in seasonal reproductive cycles, edited by B.K. Follett and D.E. Follett, Halstead Press New York: pp1-35.

Pittendrigh, C.S., Elliot, J. and Takumura, T . (1984) The circadian component in photoperiodic induction. In Photoperiodic regulation of insect and molluscan hormones, Ciba Foundation Symposium 104, edited by R. Porter and G.M. Collins, Pitman Publishing Ltd, London, pp26-47.

Pittendrigh, C.S. (1993) Temporal organization: Reflections of a Darwinian clock-watcher. Annu. Rev. Physiol. 55 : 17-54.

Pittendrigh, C.S. (1996) A letter to my friends.

Pittendrigh, C.S. and Minis, D.H. (1972) Circadian systems; longevity as a function of circadian resonance in *D. melanogaster*. Proc. Natl. Acad. Sci. USA 69 : 1537-1539.

Price, J.L., Dembinska M.E., Young M.W. and Rosbash M. (1995) Suppression of period protein abundance and circadian cycling by the *Drosophila* clock mutation *timeless*. EMBO J. 14 : 4044-4049.

Powell, J.R. (1971) Genetical polymorphism and environment. Science 176 : 545.

Ralph, M.R. and Menaker, M. (1988) A mutation of the circadian system in golden hamsters. Science 241 : 1225-1227.

Ralph, M.R., Foster, R.G., Davis, F.C. and Menaker, M. (1990) Transplanted suprachiasmatic nucleus determines circadian period. Science 247 : 975-978.

Reddy, P., Jacquier, A.C., Abovich, N., Petersen, G and Rosbash, M. (1986). The *period* locus of *D. melanogaster* codes for a proteoglycan. Cell 46 : 53-61.

Reddy, P., Zehring, W.A., Wheeler, D.A., Pirrotta, V., Hadfield, C., Hall, J.C. and Rosbash, M. (1984) Molecular analysis of the *period* locus in *Drosophila melanogaster* and identification of a transcript involved in biological rhythms. Cell 38 : 701-710.

Reppert, S.M., Tsai, T., Roca, A.L. and Surman, I. (1994) Cloning of a structural and functional homolog of the circadian clock gene *period* from the giant silkmoth *Antheraea pernyi*. Neuron 13 : 1167-1176.

Richards, R.I and Sutherlands, G.R. (1992) Dynamic mutations: A new class of mutations causing human genetic disease. Cell 70 : 709-712

Ritchie, M.G. and Kyriacou, C.P. (1994) Reproductive Isolation and the *period* Gene Of *Drosophila*. Molecular Ecology 3 : 595-599.

Roberts, D. H., Lehar, J., and Dreher, J. W. (1987) Time series with CLEAN. I Derivation of a spectrum. Journal of Astronomy 93 : 968-989

Robinson, W.P., Cambon-Thomsen,.A., Borot, N., Klitz, W. and Thomson, G. (1991) Selection, hitchhiking, disequilibrium analysis at three linked loci with application to *HLA* data. Genetics 129 : 931-948.

Rosato, E., Peixoto, A.A., Barbujani, G., Costa, R. and Kyriacou, C.P. (1994) Molecular polymorphism in the *period* gene of *Drosophila simulans*. Genetics 138 : 693-707.

Rosato, E., Peixoto, A.A., Gallippi, A., Kyriacou, C.P. and Costa, R. (1996) Mutational mechanisms, phylogeny, and evolution of a repetitive region within a clock gene of *Drosophila melanogaster*. J. Mol. Evol. 42 : 392-408.

Rosewell, K.L., Siwicki, K.K. and Wise, P.M. (1994) A *period (per)* like protein exhibits daily rhythmicity in the suprachiasmatic nuclei of the rat. Brain Research 659 : 1-236.

Royaltey,H.H, Astrahcan,E. and Sokal,R.R, (1975) Tests for patterns of geographical variation. Geographical Analysis 8 : 9-383.

Rutila, J.E., Edery, I., Hall, J.C. and Rosbash, M. (1992) The analysis of new short-period circadian-rhythm mutants suggests features of *Drosophila-melanogaster period* gene-function J. Neurogen 8 :2 : 101-113

Saez, L. and Young, M.W. (1988) *In situ* localisation of the *per* clock protein during development of *D. melanogaster* Molecular and Cellular Biology. 8 : 5378-5385.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74 : 5463-5467.

Saunders, D.S. (1982). Insect Clocks. 2nd ed. Pergamon Press, Oxford.

Saunders, D.S., Gillanders, S.W. and Lewis, R.D. (1994) Light pulse phase response curves for the locomotor activity rhythm in *period* mutants of *D. melanogaster*. J Insect Physiology 40 : 957-968.

Sehgal, A., Man, B., Price, J.L., Vosshall, L.B. and Young, M.W. (1991) New Clock Mutations In *Drosophila*. Annals Of the New York Academy Of Sciences 618 : 1-10.

Sehgal, A., Price, J.L., Man, B. and Young, M.W. (1994) Loss of circadian behavioural rhythms and *per* RNA oscillations in the *Drosophila* mutant *timeless*. Science 263 : 1603-1606.

Sehgal, A., Rothenfluh-Hilfiker, A., Hunter -Ensor, M., Chen, Y.F., Myers, M.P. and Young, M.W. (1995) Rhythmic Expression Of *timeless* - a basis for promoting circadian cycles in *period* gene autoregulation. Science 270 : 808-810.

Shin, H.S., Bargiello, T.A., Clark, B.T., Jackson, F.R. and Young, M.W. (1985) An unusual coding sequence from a *Drosophila* clock gene is conserved in vertebrates. Nature 317 : 445-448

Smith, R.F., and Konpka, R.J. (1981) Circadian clock phenotypes of chromosome aberrations with a break point at the *per* locus. Molec. Gen. Genetics 183 : 243-251.

Smith, R.F and Konopka, R.J. (1982) Effects of dosage alterations at the *per* locus on the period of the circadian clock of *Drosophila*. Mol. Gen. Genetics 185 : 30-36

Siwicki, K.K., Eastman, C., Petersen, G, Rosbash, M. and Hall, J.C. (1988) Antibodies to the *period* gene product of *Drosophila* reveal diverse tissue distribution and rhythmic changes in the visual system. Neuron 1 : 141-150.

Siwicki, K.K., Strack, S., Rosbash, M., Hall, J.C. and Jacklet, J.W. (1989) An antibody to the *Drosophila period* protein recognizes circadian pacemaker neurons in *Aplysia* and *Bulla*. Neuron 3 : 51-58.

Siwicki, K.K., Schwartz, W.J. and Hall, J.C. (1992) An antibody to the *Drosophila* period protein labels antigens in the suprachiasmatic nucleus of the rat. J Neurogen 8 : 33-42

Sokal, R.R. and Oden, N.L. (1978) Spatial autocorrelation in biology. Biological Journal of the Linnaen Society 10 199-228.

Sokal, R.R. (1979) Ecological parameters inferrred from spatial correlograms. In contemporary quantitative ecology and related ecometrics edited by G. P Patel and M. L. Rosenzweig. Maryland International Co-operative Publishing House: pp 167-196

Sokal, R.R. and Rohlf, F.J. (1981). Biometry. W.H. Freeman and Co., New York.

Sokal, R.R., Oden, N.L., Barker, J. S F . (1987) Spatial structure in *Drosophila buzzatii* populations.; simple and directional spatial autocorrelation. American Naturalist 129 : 122-142.

Strack, S . and Jacklet, J.W. (1993) Antiserum to an eye specific protein identifies photoreceptor and circadian pacemaker neuron projections in *Aplysia*. Journal of Neurobiology 24 : 552-570.

Strand, M., Prolla, T.A., Liskay, R.M. and Petes, T.D. (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365 : 274-276.

Tajima, F. (1989) Statistical Method For Testing the Neutral Mutation Hypothesis By DNA Polymorphism. Genetics 123 : 585-595.

Tautz, D. (1989) Hypervariability of simple sequences as a general source for DNA markers. Nucleic Acid Res. 17 : 6463-6471.

Thackeray, J.R. and Kyriacou ,C.P. (1990) Molecular evolution in the *Drosophila yakuba period* locus. J. Mol. Evol. 31 : 389-401

Van Gelder, R.N.,Bae, H., Palazzolo, M. and Krasnow, M.A. (1995) Extent and character of circadian gene expression in *D. melanogaster* . identification of 20 oscillating mRNAs in the fly head. Current Biology 5 : 1424-1436.

Van Gelder, R.N. and Krasnow, M.A. (1996) A novel circadianly expressed *D. melanogaster* gene dependent on the *period* gene for its rhythmic expression. EMBO Journal 15 : 1625-1631.

Van-Gelder, R.N. and Krasnow, M.A. (1996) Circadian Rhythms - Partners In Time. Current Biology 6 : 244.

Van Tol, H.M.M., Wu, C.M., Guan, H.C., Ohara, K., Bunzow,J.R., Civelli, O . and Kennedy, J. (1992) Multiple dopamine D4 receptor variants in the human population. Nature 358 : 149-152.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. and Wilson, A. (1991) African populations and the evolution of Human mitochondrial DNA. Science 253 : 1503-1507.

Vitaterna, M.H., King, D.P., Cahng, A., Kornhauser, J.M., Lowrey, P.L., McDonaled, J.D., Dove, W.F., Pinto, L.H., Turek, F.W. and Takahashi, J.S. (1994) Mutatagenesis and mapping of a mouse gene *Clock*, essential for circadian behaviour. Science 264 : 719-725.

Vosshall, L. B., Price, J.L., Sehgal, A., Saez, L. and Young, M.W. (1994) Block in nuclear localization of *period* protein by a second clock mutation, *timeless*. Science 263 : 1606-1609.

Watterson, G.A. (1975) On the number of segregating sites in genetical models without recombination. Theoretical Population Biology, 7,,256-276.

Wheeler, D.A., Kyriacou, C.P., Greenacre, M.L., Yu, Q., Rutila, J.E., Rosbash, M. and Hall, J.C. (1991) Molecular transfer of a species-specific behaviour from *Drosophila simulans* to *Drosophila melanogaster*. Science 251 : 1082-1085.

Wilks, A.V., Gibson, J.B., Oakeshott, J.G. and Chambers, G.K. (1980) An electrophoretically cryptic alcohol dehydrogenase variant in *D. melanogaster* II Post

electrophoresis heat treatment screening of natural populations. Australian Journal of Biological Sciences 33 : 575-585.

Winfree, A.T. and Gordon, H. (1977) the photosensitivity of a mutant clock. J comp. Physiol. 122 : 87-109.

Yoon, C.K. (1991) Molecular and behavioural evolution in the semispecies of *Drosophila athabasaca*. PhD Thesis Cornell Universtiy, USA.

Yoon, C.K. and Aquadro, C.K. (1994) Mitochondrial DNA variation among the *Drosophila athabasaca* semispecies and *Drosophila affinis*. Journal of Heredity 85 : 421-426.

Young, M.W. and Judd, B.H. (1978) Nonessetial sequences, genes, and polytene chromosome bands of *Drosophila melanogaster*. Genetics 88 : 723-742.

Yu, Q., Jacquier, A.C., Citri, Y., Hamblen, M., Hall, J.C. and Rosbash, M. (1987a). Molecular mapping of point mutations in the *period* gene that stop or speed up biological clocks in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 84 : 784-788.

Yu, Q., Colot, H.V., Kyriacou, C.P., Hall, J.C. and Rosbash, M. (1987b) Behaviour modification by *in vitro* mutagenesis of a variable region within the *period* gene of *Drosophila*. Nature 326 : 765-769.

Zagury, D. (1996) Between psychosis and narcissistic perversion- serial killers a clinic of horror. Journal of Evolution Psychiatrique 61 : 87-112.

Zehring, W.A., Wheeler, D.A., Reddy, P., Konopka, R.J., Kyriacou, C.P., Rosbash, M. and Hall, J.C. (1984) P-Element transformation with *period* locus DNA restores rhythmicity to mutant, arrhythmic *Drosophila melanogaster*. Cell 39 : 369-376.

Zeng, H., Hardin, P.E. and Rosbash, M. (1994) Constitutive overexpression of the *Drosophila period* protein inhibits *period* mRNA cycling. EMBO Journal 13 : 3590-3598.

Zerr, D.N., Rosbash, M., Hall, J.C. and Siwicki, K.K. (1989) Circadian rhythms of *period* protein immunoreactivity in the CNS and visual system of *Drosophila*. J. Neurosci 10 : 2749-2762

Zerr, D.M., Hall, J.C., Rosbash, M. and Siwiki, K.K. (1990) Circadian fluctuations of period protein immunoreactivity in the CNS and the visual system of *Drosophila*. J. Neuroscience 10 : 2749-2762.