Failure to reproduce *period*-dependent song cycles in *Drosophila* is due to poor automated pulse-detection and low-intensity courtship

Charalambos P. Kyriacou^{a,1}, Edward W. Green^b, Arianna Piffer^c, and Harold B. Dowse^d

^aDepartment of Genetics, University of Leicester, Leicester LE1 7RH, United Kingdom; ^bDivision of Translational Immunology, Deutsches Krebsforschungszentrum, 69120 Heidelberg, Germany; ^cMedical School, San Raffaele University, 20132 Milan, Italy; and ^dSchool of Biology and Ecology, University of Maine, Orono ME 04469-5751

Edited by Joseph S. Takahashi, Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX, and approved January 13, 2017 (received for review September 13, 2016)

Stern has criticized a body of work from several groups that have independently studied the so-called “Kyriacou and Hall” courtship song rhythms of male *Drosophila melanogaster*, claiming that these ultradian ~60-s cycles in the interpulse interval (IPI) are statistical artifacts that are not modulated by mutations at the *period* (*per*) locus [Stern DL (2014) *BMC Biol* 12:38]. We have scrutinized Stern’s raw data and observe that his automated song pulse-detection method identifies only ~50% of the IPIs found by manual (visual and acoustic) monitoring. This critical error is further compounded by Stern’s use of recordings with very little song, the large majority of which do not meet the minimal song intensity criteria which Kyriacou and Hall used in their studies. Consequently most of Stern’s recordings only contribute noise to the analyses. Of the data presented by Stern, only *per*^L and a small fraction of wild-type males sing vigorously, so we limited our reanalyses to these genotypes. We manually reexamined Stern’s raw song recordings and analyzed IPI rhythms using several independent time-series analyses. We observe that *per*^L songs show significantly longer song periods than wild-type songs, with values for both genotypes close to those found in previous studies. These *per*-dependent differences disappear when the song data are randomized. We conclude that Stern’s negative findings are artifacts of his inadequate pulse-detection methodology coupled to his use of low-intensity courtship song records.

Drosophila | courtship song | cycles | *period* gene | interpulse interval

During courtship, the *Drosophila melanogaster* male vibrates his wing toward the female and produces a series of pulses and hums (1). The pulses have a variable interpulse interval (IPI), which ranges from 15 to 80 ms but usually averages between 30 and 40 ms, whereas sympatric *Drosophila simulans* mean IPIs vary from 45 to 80 ms depending on the strain (2, 3). In 1980, in these pages, the first of a series of studies by Kyriacou, Hall, and their collaborators revealed that superimposed on these IPIs was a low-amplitude oscillation of about 60 s in *D. melanogaster* and 40 s in *D. simulans* (4–12). Furthermore, in *D. melanogaster*, these cycles were modulated in a predictable fashion by the circadian rhythm *period* (*per*) mutations (13): *per*^L males with long 29-h circadian cycles also sang with long ~80-s song cycles, whereas circadian arrhythmic *per*⁰¹ males showed a corresponding song phenotype (2, 4, 9, 10). These cycles were shown to have functional significance in playback experiments in which females were shown to be most responsive to both their species-specific IPI and cycle (2, 14–16).

The work of Kyriacou, Hall, and collaborators was performed in the late 1970s and 1980s using extremely laborious analog technology, so to attain any kind of throughput, IPIs were binned into consecutive 10-s intervals and a mean IPI calculated on a minimum of 10 IPIs (2, 4–6, 8). Initially, sine/cosine functions were fitted to the mean IPIs (2, 4–6, 8), and this was later complemented by spectral analyses, with both types of statistical methods having associated significance tests (9–12, 17). Males were raised in small single-sex groups and were usually recorded at 3 d of age with an

unreceptive <24-h virgin female, thereby prolonging the courtship. The vast majority of courtships were robust for 5–7 min, generating many hundreds of IPIs per song. Those few courtships in which one-third or more 10-s bins were empty (<10 IPIs) were not analyzed. The results from several studies (including several performed blind) revealed repeatedly and consistently that wild-type flies sang with ~50- to 65-s cycles, and that these periods were altered by *per* mutations (2, 6–8, 12).

Using different methods, Alt et al. (18) obtained similar results to the original song-cycle work, whereas Ritchie et al. (19) replicated and extended the playback experiments. It also became apparent that song cycles could be masked under conditions where the male and female were confined in small cells, when inappropriate upper IPI cut-off limits were used, or when gaps in the song record were artificially added (9, 20, 21). Against this background, Stern has used fully automated pulse detection to conclude that IPI cycles are nonexistent and consequently there are no *per*-specific differences among genotypes or species (22). We have therefore reanalyzed Stern’s primary song recordings that are provided on his website (22). Among other serious errors, we observe that Stern’s automated method only detects ~50% of song pulses. Unambiguous manual logging of pulses from Stern’s primary song records using his software platform FlySongSegmenter (23) reveals *per*-dependent song rhythms with mean period values close to those published in previous studies.

Significance

The study of ~60-s courtship song rhythms in *Drosophila* and their modulation by *period* clock mutations plays an important historical role in developing the molecular basis of the circadian oscillator. Carried out mostly in the 1980s using extremely laborious analogue methods, key features of the work were replicated by independent groups in the following decade. Recently, a study by Stern, using automated methods for song detection, has failed to reproduce these findings. By manually logging the same songs and comparing the results to the corresponding automated analyses we observe that Stern’s method detects only ~50% of the song, but in addition, incorporates spurious errors. Unambiguous manual reanalysis confirms the *period*-dependent nature of fly song rhythms as originally reported.

Author contributions: C.P.K. and H.B.D. designed research; C.P.K., E.W.G., A.P., and H.B.D. performed research; C.P.K., E.W.G., A.P., and H.B.D. analyzed data; and C.P.K., E.W.G., and H.B.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The raw data has been deposited in Dryad ([dx.doi.org/10.5061/dryad.80c1f](https://doi.org/10.5061/dryad.80c1f)).

¹To whom correspondence should be addressed. Email: cpk@leicester.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1615198114/-DCSupplemental.

Results

A Critical Assessment of Stern's Methods. Based on Stern's claims that his automated pulse-picking algorithms are highly accurate (22, 23), we initially examined the processed IPI results that Stern deposited on his website as .csv files. Each file has a series of time coordinates stretching to 45 min, with the corresponding IPI (22). We immediately noticed several problems.

Problem 1: Poor intensity courtship song. Stern analyzed all courtships exceeding 1,000 IPIs over a 45-min period, and many courtships only just exceed this threshold, giving an average of ~four IPIs per 10 s time bin (Fig. 1A). Such sporadic songs could not possibly provide any test for song cycles. The songs analyzed by Kyriacou and Hall were vigorous with few gaps, with most songs containing an average of 30–40 IPIs per 10-s bin, and if extrapolated to a 45-min recording would translate to 8,100–10,800 IPIs or 5,400–7,100 IPIs even if one-third of the bins were empty (<10 IPIs). We fortuitously found song transcripts for two genotypes processed by hand in the mid-1980s: *per⁰¹/w⁺Y* and *per⁺/w⁺Y*. The lengthier songs that did not result in copulations provided an estimate of song period (7). The average number of IPIs per 10-s bin is 33.3, predicting 8,979 IPIs in 45 min (Dataset S1). Furthermore the average proportion of gaps (<10 IPIs per 10 s) is 9.1% with 12 of 27 songs having no gaps at all (Fig. 1B and Dataset S1). These are unremarkable courtships in that most songs analyzed by Kyriacou and Hall (2, 4–12) showed this kind of vigor.

Fig. 1A shows examples of the Canton-S courtships that met Stern's criterion for analysis (>1,000 IPIs). Fig. 1A, Top shows 45 min of a Canton-S song with just over 1,000 IPIs. The IPIs have been

Fig. 1. Stern's song data. (A) Each panel shows songs analyzed by Stern (22) for 45 min, with the mean IPI calculated for every 10-s time bin. Numbers in brackets refer to the number of IPIs. Lines connect consecutive means based on 10 or more IPIs. Note that these are very sparse until the courtship has >5,000 IPIs, and even so the courtship is intermittent at the beginning of the recording period. "Start" and "Stop" on bottom panel show the points at which the reanalysis of 400 s of CS85 song analysis was initiated and ended (Dataset S2). (B) Song vigor is very low in all Stern genotypes except *per^L*. x axis: Song vigor as number of IPIs in 45 min. y axis: Proportion of songs. The vast majority of all genotypes except *per^L* generate considerably fewer than 5,000 IPIs, which was the threshold at which we analyzed songs in this study. All *per⁰¹/w⁺Y* and *per⁺/w⁺Y* songs analyzed in the mid-1980s (7) would produce >5,000 IPIs in 45 min (see also Dataset S1). (C) Mean IPIs ($\pm$ SEM) for each genotype. The arrows show the species typical IPIs for *D. melanogaster* (black) and *D. simulans* (gray). (D) The concordance between Stern's automated pulse calls compared with manual detection is poor and correlates with song vigor. The songs of the first 25 Canton-S males (including CS71; Dataset S3) were manually processed and the concordance between those IPIs correctly called by Stern and the manual method plotted against the vigor of each song (number of manually scored IPIs per second of courtship).

placed in 10-s bins as means. Only 23 points of a possible 270 are present, making any attempt to fit a cycle to these mean IPI values by spectral analysis (or any method) completely spurious, whether analyzing the complete dataset, or successive 5-min clips as performed by Stern (22), as each clip includes an average of only two to three data points. Of 312 courtships, 148 recorded by Stern had >1,000 IPIs (Fig. 1B) and were included in his analysis, yet one-third of these contained only 1,000–2,000 IPIs. Progressively more vigorous Canton-S songs are illustrated in Fig. 1A, but until songs contain >5,000 IPIs there are too many gaps for effective determination of any periodicity in the first 5–7 min, strikingly contrasting with Kyriacou and Hall's robust song production in the first few minutes of courtship (4–12) (Dataset S1).

From Stern's processed IPI records in .csv files, we were able to calculate the number of IPIs generated in the 45-min recording period for all five genotypes (22). Of Stern's 96 wild-type songs, 28 meet a criterion of >5,000 IPIs in 45 min; only 4 of 167 *D. simulans* songs reach this criterion, the vast majority have less than 1,000 (Fig. 1B); in sharp contrast, *per^L* males sing robustly with 13 of 16 having >5,000 IPIs, but only 1 of 16 of the *per⁰¹* and 2 of 16 *per^s* produce >5,000 IPIs (Fig. 1B). Consequently, to reanalyze at least a suitable fraction of Stern's data, we maintained our criterion for analysis as >5,000 IPIs in 45 min, giving us initially 28 Canton-S and 13 *per^L* songs. Two further *per^L* songs had >4,000 IPIs, so we added these to increase the sample size (Fig. 1B).

Problem 2: Inappropriate upper IPI cut-offs and poor temperature control. To distinguish an IPI from an interburst interval, Kyriacou and Hall routinely took 2 $\times$ the mean IPI as the upper IPI limit, and this was applied on a song-by-song basis (24). On inspecting Stern's *D. simulans* IPI records (22), it was clear that instead of raising the maximum IPI cut-offs to match the higher mean IPIs of this species (usually reported as ~48 ms *D. simulans* (1), Stern maintained the *D. melanogaster* cut-off of 65 ms instead of a more realistic 95 ms (22). This approach may explain why this species' mean IPI is uncharacteristically low (43.36 ± 0.38 ms, SEM) (Fig. 1C). From Stern's processed IPI files, we calculated the Canton-S mean IPI at ~35 ms, so his cut-off of 65 ms was not unreasonable, and the *per^L* mean at 37.5 ms (Fig. 1C). Kyriacou and Hall would have applied an average cut-off of 75 ms for these *per^L* songs on an individual song-by-song basis. In addition, nearly all of the *per⁰¹* and *per^s* songs gave mean IPIs >40 ms (Fig. 1C), so it would have been more appropriate to increase the IPI cut-off to ~85 ms.

We subsequently interrogated Stern's primary song Matlab files to obtain the distribution of his individual IPI calls from the songs of all five of his genotypes using a revised upper IPI cut-off of 75 ms for *per^L*, 85 ms for *per⁰¹* and *per^s*, and 95 ms for *D. simulans* and calculated the number of IPIs that would have been included using these more realistic values. We then compared these values to those obtained by Stern (22) using his exclusive 65-ms IPI cut-off (Table S1). Stern missed a tolerable 3% of *per^L* IPIs, which rose to 6% and 7% for *per^s* and *per⁰¹*, respectively. However, for *D. simulans*, Stern missed ~20% of the IPIs by using the inappropriate *D. melanogaster* cut-off (Table S1). Within Stern's Matlab files, we also found his temperature recordings for each song. The large numbers of Canton-S and *D. simulans* songs had been recorded between 23.6 and 27.9 °C, and the *per* mutants between 23.5 and 24.6 °C. Weighted mean recording temperatures for Canton-S were 25.5 °C, for *D. simulans* 25.8 °C, for *per^L* 24.6 °C, and for *per⁰¹* and *per^s* 23.8 °C. The inverse correlation between temperature and IPI (25) therefore accounts for why the latter two mutants have generally longer mean IPIs compared with *per^L* and wild type. We were surprised that this basic environmental variable, which has important implications for setting upper IPI cut-offs, was so poorly controlled.

Problem 3: Unrealistic length of courtship. Courtship interactions under natural conditions are brief, and observations in nature reveal that most interactions last less than 30 s, although a few courtships last minutes (26). Consequently, we question Stern's (22) approach of analyzing 45-min records, given that circadian cycles can damp

Fig. 3. Song rhythms of Canton-S and *perL* songs. (A) Mean periods for 10-s binned Canton-S (pink) and *perL* (blue) songs (+95% confidence limits) from cosinor, L-S, and CLEAN calculated from Stern's processed IPI files. (B) Corresponding results from manually corrected method. (C) Proportion of IPIs observed by automated compared with manual method (Stern/Manual). (D) L-S derived periods from unbinned raw IPIs from Stern's automated method compared with manually processed song pulses (20- to 150-s range). (E) Same as in D, but L-S periods from 0 to 150 s (single values of 94.7 s for CS2 and 94.5 s for *perL*10 from Stern's automated analysis, and 47.1 s for CS2 and 93.8 s for *perL*14 from the manual analyses are omitted from this figure, but not from the calculations, in order not to compress the y axis; see Dataset S2 for all results).

mean are obvious. Above the song trace in Fig. 4 is shown the frequency plot for each pulse, which is extremely helpful in determining whether the signal represents a real pulses or noise. For example, one IPI detected by Stern is generated by noise (third panel, top row in Fig. 4). This pattern of errors is reflected in the whole 300-s segment where 184 IPIs are manually logged with Stern correctly calling 48, but he also adds an additional 4 very long IPIs (but <65 ms) by pulse-skipping plus the spurious IPI produced by wing noise, giving an unacceptably low concordance rate between manual and automated analyses of 26% (48 of 184) (Fig. 4).

To pursue this further, we manually analyzed the IPIs for the first 25 wild-type songs that sang for at least 60–90 s, including our analysis of CS71. We were astonished to see that Stern's concordance rate compared with the manual analysis varied between

18% and 75%, with an average of 51.9%, but with an additional 3–4% of spurious pulses (mostly pulse skipping) (Dataset S3). We also observed a significant correlation between song vigor, measured as manually assessed IPIs generated per second of courtship and concordance ($R^2 = 0.295$, $P = 0.005$) (Fig. 1D and Dataset S3), supporting our earlier decision to select only the most vigorous songs from Stern's processed IPI files for reanalysis.

Manual Reanalysis of Stern's Vigorous Courtships Reveals period-Dependent Song Cycles in Binned and Unbinned Data. The unsatisfactory performance of Stern's automated method obliged us to manually log ~73,000 IPIs for the corresponding 25 Canton-S and 14 *perL* song segments we had earlier studied using Stern's IPI calls. From our manual analysis, which included visual (using the song trace and pulse-frequency plots) (Fig. 4) and acoustic monitoring of each pulse, we observed that Stern's automated method detected ~63% of the IPIs detected manually for both Canton-S and *perL* (Fig. 3C and Dataset S2), but this is an overestimate because we did not exclude the few percent of spurious IPIs. The observer repeated the manual analysis for each song and obtained a concordance rate of 98–99% with a few ambiguous IPIs (10–15 per song), usually at the beginning or end of a pulse train. Consequently, the intra-observer reliability of the manual method is extremely high. A second inexperienced observer reanalyzed the ~300 s of courtship song CS71 shown in Fig. 4 from the primary Matlab song recording, and detected 170 of the 184 IPIs with 10 additional putative IPIs, giving a concordance of 92.4% with the first observer. Subsequently, the first 300 s of two additional song records (CS6 and CS10) were also compared. The IPI concordance for song CS6 between the two observers improved to 97.5% (with an additional putative 1.5% IPIs detected by the naïve observer) and 99.1% for CS10 (with an additional 1.1% IPIs detected by the naïve observer). Consequently, inter- and intra-observer reliability is extremely high, and is enhanced with practice.

Fig. 2 also illustrates the corresponding manual analyses for songs CS2 and *perL*1. The arrows on the panels in Fig. 2 show that the peaks for cosinor, L-S and CLEAN are consistent for the binned data for CS2 at ~48 s, whereas for *perL*1 all peaks lie at ~145 s, supported by the correlograms (Fig. S1). In the unbinned raw data, the L-S profile for CS2 shows the peak at 47 s, whereas for *perL*1 the highest peak is in the high frequencies, yet in the 20- to 150-s range it falls at 145 s, just as in the binned data. Fig. 3B

Fig. 4. Stern's automated pulse detection is highly insensitive compared with manual analysis and also generates IPI artifacts. The first 300 s of recording for song CS71 are shown in FlySongSegmenter. Stern's pulse calls in blue, manual pulse calls in red. Seven examples of pulse skipping (green line) by the automated method are shown that create very long IPIs, four of which (<65 ms) were included in Stern's analysis, as was one spurious IPI (third song panel, top row). The Upper panel in each row shows the pulse frequency plot.

summarizes the cosinor, L-S, and CLEAN results of binned data from the corresponding manually corrected song files. Statistical analysis revealed a highly significant genotype effect ($F_{1, 111} = 31.2$, $P \sim 0$), with no significant effects for the different statistical methods nor any interaction. The mean period from cosinor of Canton-S songs taken from the binned data were 60.8 s and that of *per^L* was 87.6 s (Fig. 3B). With the L-S and CLEAN algorithms, Canton-S songs vs. *per^L* gave mean periods of 50.1 s vs. 83.1 s and 48.1 vs. 72.9 s, respectively (Fig. 3B). We confirmed all these results with the Mann–Whitney *U* test (cosinor $P = 0.005$, L-S $P = 0.007$, CLEAN $P = 0.009$; all are significant with Bonferroni corrections, even when two-tailed). Furthermore, when we initiated our analysis 5 s through the first bin to generate a maximal displacement of IPIs from their original 10-s bins, cosinor analysis still revealed a significant difference in the periods between the two genotypes (49.6 vs. 74.5 s, $F_{1, 37} = 6.8$, $P = 0.013$ or $P = 0.032$ by Mann–Whitney *U* test, two tailed). Consequently, cosinor analysis does not appear to be particularly sensitive to the IPI position within the bin at which the analysis is initiated. We also examined the raw unbinned IPI data using L-S, and again the high frequencies of 0.8–5 s dominated, with no significant differences between the two genotypes (Fig. 3E, Right) (Mann–Whitney *U* test, $P = 0.59$). When we took the highest peak in the 20- to 150-s range, we observed a mean period of 46.1 s for Canton-S and 72.2 s for *per^L* (Fig. 3D, Right) ($F_{1, 37} = 8.59$, $P = 0.006$, Mann–Whitney *U* test, $P = 0.009$, two tailed).

period-Dependent Differences in Song Cycles Disappear in Randomized IPI Datasets. We then randomized both the binned and corresponding raw IPI datasets, maintaining Stern's interburst intervals of >65 ms for each song and analyzed with L-S. We compared the distributions of the observed to the corresponding randomized binned data by placing the periods into four categories reflecting the *per⁰* (20–30 s), *per^S* (30–45 s), wild-type, (45–70 s), and *per^L* (70–150 s) domains (9), and performed a goodness-of-fit test. For both the binned wild-type and *per^L* data, the distributions of the observed and randomized binned periods were highly significantly different ($\chi^2 = 20.9$, $P = 0.0001$ for Canton-S with infinite χ^2 for *per^L* as none of the randomized data period fell into the *per^L* domain, in contrast to 10 of 14 of the observed *per^L* periods). A test of independence comparing the observed Canton-S with *per^L* period distributions was also highly significant, as expected ($\chi^2 = 19.6$, $P = 0.0002$).

For the unbinned IPIs, the peak periods, nearly always in the high frequencies, were placed into four categories corresponding to periods of <1.0, 1–2, 2–5, and >5 s. The goodness-of-fit tests generated significant deviations of the observed from the randomized data (Canton-S $\chi^2 = 26.4$, $P \sim 0$, *per^L* $\chi^2 = 21.3$, $P < 0.0001$). A test of independence of the observed Canton-S and *per^L* distributions gave a marginal difference ($\chi^2 = 8.7$, $P = 0.034$). When we repeated the goodness-of-fit test for the peak periods in the 20- to 150-s domain, there was no significant difference between the observed and randomized Canton-S distributions ($\chi^2 = 3.0$, $P = 0.39$) in contrast to *per^L* ($\chi^2 = 11.2$, $P = 0.011$). A test of independence of the observed 20- to 150-s periods from unbinned data between the two genotypes was also significant ($\chi^2 = 9.91$, $P = 0.019$), further confirming the significance of the period difference between them in this temporal domain.

Stern's Automated Method Performs Poorly Even When the Signal-to-Noise Ratio Is Improved. Finally, we wondered whether the poor performance of Stern's automated pulse detection was also observed in his initial paper, in which he describes the FlySongSegmenter methodology (23). In that study, Stern used much smaller mating cells that focused the flies closer to the sensitive part of the microphone and, consequently, the song pulses would be expected to be louder and generate a higher signal-to-noise ratio, even though under such confined conditions song rhythms are masked

(9). We took the 16 Canton-S Tully song records from that work (available as *.wav files) (23) and analyzed up to a few hundred pulses from each song manually, then compared these results to his automated method. We observed an average concordance of 75%, with the best song having an impressive 96% but the worst 11% (Table S2). Furthermore, each song had a number of IPIs automatically detected that were spurious. For example, the song that had the 11% concordance rate had only detected 65 of 329 IPIs but of these 65, 29 were spurious, so only 36 of the IPIs had been successfully detected, generating the concordance value of 11% (36 of 329). The song with the 96% concordance was the only one in which Stern's automated method selected more IPIs than the manual method (111 vs. 93). However, the IPI analysis is severely contaminated because 22 of the 111 pulses were spurious.

Stern's own comparison of his automated and manual methods using 1 min of song from each of nine wild-type songs shows a mean predictive value, *F*, of ~ 0.9 (23). Our more pragmatic analysis of nearly 66 min of song from 16 wild-type flies from this study suggests a far less-impressive performance for the automated method. Consequently, when using smaller mating cells, the mean concordance rate is enhanced to 75%, but so is the spurious IPI rate to 7% (Table S2), so any improvement in sensitivity using smaller mating cells is largely offset by the higher rate of incorrect pulse detection.

Discussion

Serious methodological flaws have been uncovered in Stern's song analyses (22). We are at a loss to explain why Stern's *per^L* songs are as vigorous as those studied by Kyriacou and Hall (2, 4–12), whereas all of the other genotypes sing so poorly. Even if we assume Stern missed one-third to one-half of the IPIs of *per⁰* and *per^S*, these males would not sing nearly enough to be included in any song rhythm analysis using Kyriacou and Hall criteria (Fig. 1C). The use of courtships with very sparse song production adds a major noise component to Stern's results. The lack of temperature control in Stern's study is startling, given the well-known dependence of IPI on temperature (25). Equally puzzling is the adoption of *D. melanogaster* upper IPI limits when analyzing *D. simulans* songs. However, Stern's apparent reluctance to manually monitor extended samples of his songs in his latter study is astonishing, as he would have immediately observed a concordance rate of only $\sim 50\%$. On the basis of his critically flawed analysis, Stern has sought to discredit a decade's worth of results on song cycles from Kyriacou, Hall, and collaborators (2, 4–12), as well as confirmatory work from other independent groups (18, 19). From our exhaustive manual reanalyses of nearly 6 h of Stern's courtship song records, we seldom observed consistently reliable automated scoring for any 10-s segment of song from his study (22) (Fig. 4).

We have documented that when the most vigorous of Stern's song records are analyzed manually and IPIs placed into 10-s bins, *per^L* songs have significantly longer periods than Canton-S using three different time-series analyses. The distribution of periods for each genotype is also significantly different from the corresponding randomized data. For unbinned data, we concur with Stern that the highest peaks in the L-S periodogram are nearly always in the very short period range of 0.8–2 s, and these show a very marginal difference between the genotypes. However, the period distributions of these high-frequency cycles are not random. This may be caused by the general increase in IPI reported to occur during a burst (35), a pattern that would generate high frequencies in the L-S periodogram and be further modulated by interburst intervals. When we examined the periods of the unbinned raw IPIs within the 20- to 150-s domain, we again see significantly longer periods in *per^L* compared with wild type, which is also supported by comparison of the corresponding period distributions of the two genotypes. Given that each male sings such a broad range of IPI lengths ranging from 15 to 65 ms, it seems unlikely that an individual species-specific IPI value acts as a trigger to release

female mating, as originally envisioned (1). It seems more likely that females average IPIs over short periods of time. If each female has a slightly different preferred mean IPI, then the cycling of IPI could act to allow the male to scan the IPI preferences of different potential partners as they court them successively (36). Consequently, mean IPIs taken over short periods of time could represent more biologically relevant stimuli than individual IPIs.

In terms of mechanism, the effects of *per* on ~60-s song rhythms cannot be explained by the canonical circadian transcription–translation loop. However, ultradian (~30 min) hormone rhythms are also known to be altered by circadian clock mutations in hamsters (37), as is the timing of critical period plasticity in cortical neurons during mouse postnatal maturation (38). The fly song cycle may be a pleiotropic effect of *per*, which acts on thoracic ganglia during development, possibly through *per*-expressing glia (28).

Our reinspection of Stern's data and his methods suggests that considerable caution needs to be exercised in accepting conclusions from song studies using the automated pulse detection within FlySongSegmenter (23). Although it might be acceptable to calculate an average IPI value for a complete courtship based on, for example, 2,000 rather than 3,000 IPIs (where it is unlikely to make much overall difference), omitting significant proportions of the data and adding spurious IPIs can generate major problems on 10-s or finer time scales. This problem particularly applies to studies where much more sensitive and dynamic song features are

extracted using FlySongSegmenter. For example, changes reported by Stern to occur over the length of a song burst may need to be revised, given that we have documented how discrete and complete song bursts cannot be reliably defined because of undetected pulses (39). Furthermore, quantifying pulse-song intensity is also unlikely to be reliable, because the automated software ignores so much low-intensity pulse song. Indeed any measure that includes pulse song as a variable needs to be reexamined manually (40–42).

In summary, we have demonstrated that Stern's automated pulse detection method is critically flawed (22). Manual reanalysis of his primary songs records reveal IPI periods of ~80 s in *per*^L and 50–60 s in wild type. We acknowledge that despite our attempts to be as accurate as possible, some human error will inevitably creep into such a comprehensive manual analysis. However, these errors pale into insignificance when compared with the systematic omissions and spurious IPIs generated by Stern's automation. We appreciate that automated methods can never be perfect, but Stern appears to have severely overstretched the trade-off between convenience and accuracy.

Methods

Further details can be found in [SI Methods](#).

ACKNOWLEDGMENTS. We thank Dr. D. Heslop for kindly providing us with a working version of CLEAN.

- Ewing AW, Bennett-Clark HC (1968) The courtship songs of *Drosophila*. *Behaviour* 31: 288–301.
- Kyriacou CP, Hall JC (1986) Interspecific genetic control of courtship song production and reception in *Drosophila*. *Science* 232(4749):494–497.
- Kawanishi M, Watanabe TK (1980) Genetic variations of courtship song of *Drosophila melanogaster* and *Drosophila simulans*. *Jpn J Genet* 55:235–240.
- Kyriacou CP, Hall JC (1980) Circadian rhythm mutations in *Drosophila melanogaster* affect short-term fluctuations in the male's courtship song. *Proc Natl Acad Sci USA* 77(11):6729–6733.
- Kyriacou CP, Hall JC (1985) Action potential mutations stop a biological clock in *Drosophila*. *Nature* 314(6007):171–173.
- Zehring WA, et al. (1984) P-element transformation with *period* locus DNA restores rhythmicity to mutant, arrhythmic *Drosophila melanogaster*. *Cell* 39(2 Pt 1):369–376.
- Hamblen M, et al. (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(5): 249–291.
- Yu Q, Colot HV, Kyriacou CP, Hall JC, Rosbash M (1987) Behaviour modification by *in vitro* mutagenesis of a variable region within the *period* gene of *Drosophila*. *Nature* 326(6115):765–769.
- Kyriacou CP, Hall JC (1989) Spectral analysis of *Drosophila* courtship song rhythms. *Anim Behav* 37:850–859.
- Kyriacou CP, van den Berg MJ, Hall JC (1990) *Drosophila* courtship song cycles in normal and *period* mutant males revisited. *Behav Genet* 20(5):617–644.
- Dushay MS, et al. (1990) Phenotypic and genetic analysis of *Clock*, a new circadian rhythm mutant in *Drosophila melanogaster*. *Genetics* 125(3):557–578.
- Wheeler DA, et al. (1991) Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster*. *Science* 251(4997):1082–1085.
- Konopka RJ, Benzer S (1971) Clock mutants of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 68(9):2112–2116.
- Kyriacou CP, Hall JC (1982) The function of courtship song rhythms in *Drosophila*. *Anim Behav* 30:794–801.
- Griffith LC, et al. (1993) Inhibition of calcium/calmodulin-dependent protein kinase in *Drosophila* disrupts behavioral plasticity. *Neuron* 10(3):501–509.
- Greenacre ML, Ritchie MG, Byrne BC, Kyriacou CP (1993) Female song preference and the *period* gene in *Drosophila*. *Behav Genet* 23(1):85–90.
- Demetriades MC, Thackeray JR, Kyriacou CP (1999) Courtship song rhythms in *Drosophila yakuba*. *Anim Behav* 57(2):379–386.
- Alt S, Ringo J, Taly B, Bray W, Dowse H (1998) The *period* gene controls courtship song cycles in *Drosophila melanogaster*. *Anim Behav* 56(1):87–97.
- Ritchie MG, Halsey EJ, Gleason JM (1999) *Drosophila* song as a species-specific mating signal and the behavioural importance of Kyriacou & Hall cycles in *D. melanogaster* song. *Anim Behav* 58(3):649–657.
- Ewing AW (1988) Cycles in the courtship song of male *Drosophila melanogaster* have not been detected. *Anim Behav* 26:1091–1097.
- Crossley SA (1988) Failure to confirm rhythms in *Drosophila* courtship song. *Anim Behav* 36:1098–1109.
- Stern DL (2014) Reported *Drosophila* courtship song rhythms are artifacts of data analysis. *BMC Biol* 12(1):38.
- Arthur BJ, Sunayama-Morita T, Coen P, Murthy M, Stern DL (2013) Multi-channel acoustic recording and automated analysis of *Drosophila* courtship songs. *BMC Biol* 11(1):11.
- Hall JC, Kyriacou CP (1990) Genetics of biological rhythms in *Drosophila*. *Adv Insect Physiol* 22:221–298.
- Shorey HH (1962) Nature of the sound produced by *Drosophila melanogaster* during courtship. *Science* 137(3531):677–678.
- Gromko MH, Markow TA (1993) Courtship and remating in field populations of *Drosophila*. *Anim Behav* 45:253–262.
- Yamazaki S, Takahashi JS (2005) Real-time luminescence reporting of circadian gene expression in mammals. *Methods Enzymol* 393:288–301.
- Konopka RJ, Kyriacou CP, Hall JC (1996) Mosaic analysis in the *Drosophila* CNS of circadian and courtship-song rhythms affected by a *period* clock mutation. *J Neurogenet* 11(1–2):117–139.
- Refinetti R, Lissen GC, Halberg F (2007) Procedures for numerical analysis of circadian rhythms. *Biol Rhythm Res* 38(4):275–325.
- Roberts DH, Lehar J, Dreher JW (1987) Time series with CLEAN. I. Derivation of a spectrum. *Astron J* 93:968–989.
- Chatfield C (1996) *The Analysis of Time Series* (Chapman and Hall, London), 5th Ed.
- Dowse HB (2009) Analyses for physiological and behavioral rhythmicity. *Methods Enzymol* 454:141–174.
- Logan IG, Rosenberg J (1989) A referee's comments on the identification of cycles in the courtship songs of *Drosophila melanogaster*. *Anim Behav* 37:860.
- Heslop D, Dekkers MJ (2002) Spectral analysis of unevenly spaced climatic time series using CLEAN: Signal recovery and derivation of significance levels using a Monte Carlo simulation. *Phys Earth Planet Inter* 130:103–116.
- Ewing AW (1983) Functional aspects of *Drosophila* courtship. *Biol Bull* 58:275–291.
- Kyriacou CP, Ritchie M, Greenacre ML (1992) Genetics of sexual selection in *Drosophila*. *Am Zool* 32:31–39.
- Loudon AS, et al. (1994) Ultradian endocrine rhythms are altered by a circadian mutation in the Syrian hamster. *Endocrinology* 135(2):712–718.
- Kobayashi Y, Ye Z, Hensch TK (2015) Clock genes control cortical critical period timing. *Neuron* 86(1):264–275.
- Coen P, et al. (2014) Dynamic sensory cues shape song structure in *Drosophila*. *Nature* 507(7491):233–237.
- Clemens J, et al. (2015) Connecting neural codes with behavior in the auditory system of *Drosophila*. *Neuron* 87(6):1332–1343.
- Coen P, Xie M, Clemens J, Murthy M (2016) Sensorimotor transformations underlying variability in song intensity during *Drosophila* courtship. *Neuron* 89(3):629–644.
- Shirangi TR, Wong AM, Truman JW, Stern DL (2016) Doublesex regulates the connectivity of a neural circuit controlling *Drosophila* male courtship song. *Dev Cell* 37(6): 533–544.
- Dowse HB, Ringo JM (1987) Further evidence that the circadian clock in *Drosophila* is a population of coupled ultradian oscillators. *J Biol Rhythms* 2(1):65–76.
- Kay SM, Marple SGJ (1981) Spectrum analysis, a modern perspective. *IEEE Proc* 69: 1380–1419.
- Bretthorst GL (2001) Nonuniform sampling: Bandwidth and aliasing. *Bayesian Interference and Maximum Entropy Methods in Science and Engineering*, eds Rychert J, Erickson G, Smith CR (Kluwer, Dordrecht, The Netherlands), pp 1–28.