Title

A macroevolutionary gene network reveals diapause evolutionary dynamics beyond the circadian clock and predicts microevolution

Saurav Baral[1], Sridhar Halali[2], Mats Ittonen[1,3], Klaus Fisher[4], Karl Gotthard[1], and Christopher Wheat[1]*

[1] Stockholm University, Department of Zoology, 10691 Stockholm, Sweden

[2] Biodiversity & Evolution Unit, Dept. of Biology, Lund University, Sweden

[3] Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia

[4] Institute for integrated natural Sciences, department of Biology, Zoology, University of Koblenz-landau, Koblenz, Germany

Abstract

Diapause is an alternative developmental pathway evolved independently in many insects to synchronize life cycles with resource abundance. While subsets of this essential phenotype have long been studied at a single species level, the genomic basis of the full diapause syndrome remains poorly understood. Remaining unknown is whether convergent diapause syndromes employ shared mechanisms. This paucity of insights has fueled a long-standing debate about how life cycle synchronization evolves. Using a comparative genomic analysis spanning diverse diapause transitions in butterflies, we identified a large network of coevolving genes unique to diapausing species. The network is composed of functional modules spanning circadian regulation, metabolism, and cell cycle control. We tested whether this macroevolutionary scale network predicts microevolutionary dynamics, hypothesizing that this network is the polygenic architecture underlying the diapause syndrome. Analyses revealed that allelic variation in the diapause network is significantly enriched in signatures of local adaptation across latitudes, but only in diapausing species. Thus, we empirically show that diapause evolves through modular, coevolving gene networks, components of which regulate species/population specific diapause phenotypes. This novel perspective on this complex phenotype opens a new horizon of inquiry for understanding seasonal adaptation across evolutionary scales, while demonstrating the power of using comparative genomics to dissect polygenic phenotypes.

Keywords

diapause, comparative genomics, butterflies, circadian clock, coevolutionary gene network, macro-to-micro evolution

Introduction (3000 words)

Many temperate insects synchronize their life cycles with seasonal variation in resource abundance using an environmentally-induced dormancy called diapause(*1*, *2*). Diapause is a state of low physiological activity involving metabolic suppression, developmental arrest, and increased stress resistance, enabling survival until favorable conditions return(*3–5*). Understanding the genetic basis of diapause and how it evolves is a long sought-after goal, with a "diapause toolkit" used across species often envisioned(*6*), but whether such a coherent set of genes exists has been called into question(*7*). Nevertheless, mechanistic insights into diapause hold the potential to reveal how a key innovation in life-cycle regulation has evolved, which can inform modeling of insect population responses to diverse anthropogenic challenges(*1*, *3*, *8*, *9*).

Diapause is a syndrome(*4*), integrating a wide array of phenotypes and trade-offs: arresting development, slowing biological time, enhancing cellular homeostasis, and increasing resistance to stress, often resulting in delayed maturation, reduced reproductive investment, and extended lifespan. There are likely many different ways to achieve this alternative life-history pathway, leading some researchers to expect that rather than individual genes being shared across independently evolved diapause phenotypes, functional pathways themselves might be a more coherent analysis framework(*7*). This complexity of the diapause syndrome challenges mechanistic dissection of its origin and evolutionary dynamics due to (a) its polygenic nature(*7*, *10)* and (b) a recognized gap in evolutionary biology, the micro-to-macro evolutionary gap(*11*). Syndromes, like diapause and other complex phenotypes, are readily observed on a macroevolutionary scale, yet these innovations often lack relevant trait variation for the full syndrome within species(*12*). Without relevant microevolutionary variation, establishing genotype to phenotype connections is severely constrained(*6*). Hence, while dissection of the intra-specific variation can help build models of what is required for subsets of syndrome phenotypes, the ability of such insights to inform upon the origins and evolutionary dynamics of syndromes as a whole is limited. Thus, a new approach is needed to study the evolution of the diapause syndrome.

These challenges are well exemplified in Lepidoptera, wherein numerous population genomic, physiological, and transcriptomic studies have sought mechanistic insights into diapause traits. While genomic comparisons between populations (or strains) differing in diapause induction thresholds often find strong evidence for a role of alleles in circadian clock genes(*13*, *14*), diapause physiology via transcriptomic studies consistently reveal diverse functional pathways associated with the developmental regulation of diapause(*15*). Despite this impressive progress, remaining unknown are the genetic mechanisms influencing diapause evolution beyond its induction, and whether these mechanisms are shared across diverse species (*6*, *7*). Further, while diapause is known to have evolved independently many times and manifests at different life stages (egg, larva, pupa, or adult) across insect species(*4*, *5*), generally unknown across insects is whether species within the same family have a shared or independent origin of diapause.

Comparative analysis of diapause

Here we address these aforementioned challenges using recent advances in both comparative genomics(*16*) and detailed reconstructions of diapause syndrome evolution in butterflies(*17*, *18*). Butterflies (superfamily Papilionoidea) now have more than 100 high-quality genomes available spanning nearly 100 MY of evolution and well-defined ecological traits(*17–20*). This combination offers a unique and powerful framework wherein comparative genomic analyses can be used to

bridge the micro-to-macro evolutionary gap to reveal the genetic mechanisms underlying complex syndromes, such as diapause in insects.

We began by gathering high-quality genomes from species with known diapause phenotypes and assessing correlations between the evolutionary rate of protein-coding genes with presence/absence of diapause. Our dataset includes 89 species from four major families spanning nearly 100 MY of evolution (Fig. 1, A and B; table S1). Diapause in these species spans ~16 independent evolutionary transitions of the diapause syndrome, which manifests across four different life stages (Fig. 1B;(*17*). We approach this complexity by using a binary encoding of species having or lacking diapause, searching for general insights into diapause mechanisms.

Evolutionary rates of diverse genes are correlated with diapause

In order to gain general insights into diapause evolution, our sampling timescale spans at least 100 million generations. This covers a vastly deeper temporal horizon than previous large-scale comparative genomics projects (e.g., while taxonomic sampling in the mammalian-focused Zoonomia project spanned nearly 100 MY of evolution, those species' generation times were >> 1 year)(*16*). At this evolutionary scale, alignments of putatively regulatory regions fail to pass rigorous quality filtering(*21*), limiting our analysis to a total of 4,571 single-copy orthologs that could be aligned with high confidence. These were used to assess whether any of these genes have evolved at different rates along branches leading to species with and without diapause (while accounting for rate variation among genes as well as phylogenetic non-independence). For this we calculated Rho, which measures the correlation between each gene's relative evolutionary rate in relation to the presence or absence of diapause tip traits. Here, evolutionary rate estimation uses the total amount of genetic change (i.e. analyses are not based upon dN or dS values). Genes with significant positive Rho values indicate increased evolutionary rate in diapausing lineages (Fig. 2A; table S7). In contrast, genes with significant negative Rho values evolve slower in diapausing lineages(*16*, *22*, *23*).

A total of 59 genes have evolutionary rates significantly correlated with diapause (q-value < 0.1; $p < 0.0017$; Fig. 2A), with 25 having a positive association (Rho > 0), while the rest are negatively associated (Rho < 0). Notably, four of the 25 outliers showing strong positive evolutionary rates associated with diapause are the core circadian clock genes *period* (*per*) (Rho = 0.30, p-value = 0.00001), *timeless (tim)* (Rho = 0.25, p-value = 0.0015), *clock* (Rho = 0.28, p-value = 0.0007), and *cycle* (Rho = 0.24, p-value = 0.0009). These macroevolutionary results are strikingly concordant with diverse microevolutionary-level studies that commonly report associations and causal connections between allelic variation in these clock genes and diapause induction phenotypes across diverse Lepidoptera and other insects(*14*, *24*, *25*). Beyond providing strong validation for our comparative analysis, finding that the evolutionary rate of multiple clock genes is positively associated with diapause formally integrates diverse species level insights, suggesting that the clock genes are best viewed as parts of an interconnected gene network. We next sought to further explore these significant Rho genes to gain mechanistic insights into the diapause syndrome.

Gene set enrichment analyses (GSEA) using gene ontology (GO) terms on the positive Rho genes (n = 25) revealed an enrichment for circadian processes, while negative Rho genes (n = 34) were enriched for metabolism and development functions (Fig. 2C). After removing the 4 clock genes (*per, tim, clock and cycle*), GSEA on the remaining positive Rho genes (n=21) revealed an enrichment of negative regulators of tissue development (fig. S4). Using functional annotations from *Drosophila melanogaster*, these 25 positive Rho genes fall into seven subjective groups: circadian regulation/clock genes (4 genes), chromatin regulation (4 genes), membrane receptors

and signaling (3 genes), membrane and vesicle transport (4 genes), core metabolism (4 genes), cytoskeletal and developmental morphogenesis (4 genes), and protein and DNA maintenance (2 genes) (table S12). These non-clock, positive Rho genes associated with diapause evolution represent the first glimpse of previously unknown genes and genetic pathways potentially impacting diapause evolution, demonstrating a route beyond a circadian clock focus. However, like many in the non-model genomics community, we find GO term analyses and functional annotations inferred across deep evolutionary time to be of limited value for *post-hoc* functional insights (especially when annotations primarily rely upon species lacking a coherent diapause phenotype). Given the power of our approach to uncover both known and novel genes associated with the diapause syndrome beyond circadian regulation, and the limitations of GO term analyses, we next explored alternative routes to study the genetic mechanisms of diapause evolution.

Evolutionary Rate Correlations

Genes encode products that have intimate functional interactions with other loci (e.g., via protein-protein, metabolic pathway or gene regulatory network connectivity). Because genes evolve while maintaining these interactions, genes coevolve with their interacting partners. Macroevolutionary analysis of correlations in evolutionary rates between genes can reveal such co-evolutionary networks and provides a powerful means of predicting gene function(*26*, *27*). Measures of evolutionary rate covariation (ERC) estimate the strength of correlation in evolutionary rates between two genes, across all the branches of a phylogenetic tree (while correcting for phylogenetic relatedness). High ERC values indicate co-evolving genes, suggesting either functional interactions, co-regulation, physical interactions, or shared evolutionary constraints(*27*).

Using our dataset of evolutionary rates for 4,571 genes (Fig. 2), we computed 10,444,735 pairwise ERC values across three dataset partitions (Fig. 3F): the full species set (n = 89), only diapausing species (n = 54), only non-diapausing species (n = 35). Circadian clock genes were used to verify and calibrate our ERC results, as their protein-protein and regulatory interactions are well documented and conserved across Bilateria (fig. S11) (*28*, *29*). Components of the circadian clock that physically interact (*period*, *timeless*, *cry2*, *clock* and *cycle*; Fig. 3, D, G, H) exhibit strong co-evolutionary dynamics (ERC values > 0.4;), while those with only regulatory interactions are weaker (Fig. 3, E, G, H). Going forward, to discover the gene networks potentially underlying diapause evolution, we used this empirically derived threshold between genes (ERC > 0.4) as our co-evolving cutoff, which represents the top 0.52 % of all ERC values (Fig. 3F, red line; fig. S12).

Evolutionary dynamics of the circadian clock in diapausing lineages

Whether circadian clock genes might be evolving differently in diapausing vs. non-diapausing lineages has long been implied in the literature but never formally quantified. We find that specific components of the circadian clock exhibit stronger coevolution in diapausing lineages. For example, in the diapausing species set, the ERC for *period* - *timeless* is 0.66, while in the non-diapausing set this is weaker (0.38). In the full dataset of all species (n = 89), the *period* - *timeless* ERC is intermediate between the previous values (0.60). Extending ERC analyses across all the core circadian clock genes (*period*, *timeless*, *cry1*, *cry2*, *clock* and *cycle*) reveals a strong interconnected network of coevolution in diapausing lineages not seen in non-diapausing lineages (Fig. 3I). Thus, different evolutionary dynamics are acting upon these core clock genes when evolving in lineages with vs. without diapause. This insight advances the debate regarding to what extent the evolutionary dynamics of the circadian clock are associated with seasonality(*30*). Our

results support the hypothesis that core components of the circadian clock are part of a coevolving network of diapause genes.

Evolutionary dynamics beyond the clock: the Rho-gene network

Given the power of ERC analysis to detect expected coevolutionary dynamics (e.g., the interacting circadian clock genes; Fig. 3H) and provide novel insights (Fig. 3I), analyses were extended to identify the full set of coevolving genes associated with the 25 genes significantly associated with diapause (positive Rho genes, Fig. 2B). The network of coevolving genes in diapausing lineages comprises two sets of 14 and 2 genes each, with core clock genes being part of the larger set (Fig. 4A). Non-diapausing lineages lack such a large co-evolving network, having only two small sets of 4 and 3 genes, with clock genes comprising the smaller set (just *period, clock* and *cycle*; Fig 4A). The genes comprising the large coevolutionary network in diapausing lineages (n=14 genes) are hereafter referred to as the "Rho-gene network". The emergence of Rho-gene network from our macroevolutionary analysis is intriguing, as it suggests evolutionary dynamics in diapausing lineages that include and extend beyond clock genes.

Using our Rho-gene network to investigate diapause evolution allows moving beyond candidate genes and towards a formal network level analysis that captures the breadth of genes involved in the full diapause syndrome. While members of the Rho-gene network are significantly associated with the evolution of diapause (Fig. 2B), each of these genes is also potentially coevolving with a larger set of genes. Discovering this larger network holds potential for studying the evolution of a complex phenotype like the diapause syndrome. If selection on the diapause syndrome during evolution acts primarily at the network level, selection on individual genes may be diffuse and the contribution of genes within the network variable across species(*7*). Such diffuse signals are difficult to detect on a microevolutionary scale, challenging to reconcile across independent species-level studies, and generally hampered by gene functional annotations that lack gene set grouping for the trait of interest. To formally address these challenges, we use the identified Rho-gene network to develop a formal hypotheses framework for studying the evolution of diapause, deriving a macroevolutionary informed gene set for *a priori* testing that is fundamentally different from the *post-hoc* analyses of gene ontology term groupings in GSEAs.

Starting with the Rho-gene network (n = 14 genes), additional genes were retained from the full dataset (n=4,557 genes) if they exhibited strong co-evolutionary signatures (ERC > 0.4 in the diapausing taxa subset) with any of the Rho-gene network members. This yielded a coevolving network of 504 genes, hereafter referred to as the "expanded Rho-gene network". Gene ontology-based enrichment analysis was largely unhelpful in revealing the biological function of this larger network (fig. S3). Instead, we took a network-based approach based on ERC values to detect modularity within the expanded Rho-gene network (Fig. 4B). Stochastic block model-based clustering identified eleven modules (table S10) within the expanded Rho-gene network. Each of these modules is enriched in genes that belong to different biological processes (fig. S8). For example, Cluster 8 (n = 29 genes) is enriched for circadian processes (*period*, *timeless*, *clock*, etc.), while Cluster 10 (n = 26 genes) is enriched in cell division and cell cycle regulators. Taken together, these coevolutionary modules span the different physiological aspects of the diapause syndrome, potentially revealing the modular architecture through which diapause evolves.

A putative diapause syndrome "toolkit" would be expected to include processes such as cell cycle arrest, metabolic suppression, and heightened stress and defense responses, reflecting themes that recur across species(*7*). Comparative work across a broader range of species within and outside Papilionoidea will be critical to determine whether the gene clusters and pathways identified here

represent general mechanisms of diapause evolution. Notably, several candidate genes previously implicated in diapause in taxa beyond Lepidoptera were absent from our expanded Rho network (table S14). While this absence could reflect butterfly-specific features of diapause regulation, it more likely indicates that shared biological processes, rather than identical genes, underlie diapause across lineages. Similar conclusions have emerged from diverse transcriptomic studies, which show that pathway-level rather than gene-level conservation may better characterize the evolution of diapause(*6*, *7*). These insights underscore the importance of adopting a network framework for comparative assessments of the diapause syndrome, in which species may rely on different components of interconnected gene clusters to achieve similar phenotypic outcomes. Such a perspective cautions against assuming that candidate genes identified in one lineage will apply universally and highlights the value of investigating trait evolution through the lens of co-evolving gene networks rather than isolated loci.

Microevolution

To further explore the utility of our network framework for understanding the convergent evolution of the diapause syndrome, we tested whether the expanded Rho-gene network, identified at a macroevolutionary level, is enriched in signatures of local adaptation arising from microevolutionary dynamics across diverse butterfly species. For this, we used population genomic analyses of geographically distinct populations in three species with, and one without diapause (Fig. 4C). Diapause phenotypes have either been shown to differ between these populations (*31*, *32*) or on similar geographic scales (*33*, *34*). We then tested whether there was a significant enrichment of expanded Rho-gene network (n= 504) in genomic regions exhibiting signatures of local adaptation (top 5% F_{ST} outlier loci; similar results using top 1% outliers are in fig. S16). Genes in the expanded Rho-gene network were significantly enriched in three species with diapause compared to 1000 random sets of 504 genes and distributed across the whole genome (Fig. 4C), with roughly one-third (34%; n= 95) common across the three diapausing species (Fig. 4D). However, no such enrichment was seen in our non-diapausing species (*Papilio polytes*). The intersection genes from diapausing species are distributed across the eleven modules produced by the SBM clustering of the expanded Rho-gene network (fig. S10). These results not only provide empirical support for our macroevolutionary predictions, but also reinforce the view that the diapause syndrome has a modular genetic architecture, shaped by ongoing microevolutionary dynamics that give rise to the coevolving gene modules of the diapause syndrome.

Conclusion/Discussion

While species-level studies have made great strides in discovering mechanisms of diapause functioning, these advances primarily derive from candidate gene knockout studies and investigations of large effect alleles, which often differ across taxa (table S15) (*6*). Attempts to extrapolate from these mechanistic findings to general evolutionary insights have been challenging (*1*, *30*), exemplified by the ongoing debates about the relative role of circadian clock components, their functioning, and the possible role of additional physiological or metabolic mechanisms' contribution to diapause (*35–37*). Here we have taken a fundamentally different integrative approach that bridges macro- and microevolutionary scales. We identified a clade-specific gene coevolutionary network that is independent of functional annotations, enabling formal hypothesis testing at both the macro- and microevolutionary levels. This design allowed us to move beyond *post hoc* gene set comparisons via GSEA to test *a priori* predictions directly, providing a powerful

and generalizable framework for understanding how complex adaptive phenotypes evolve. In doing so, we establish a quantitative and empirical connection between deep evolutionary history and contemporary genetic variation, something long called for but rarely achieved in evolutionary biology.

We posit that the diapause syndrome has evolved via a network of coevolving genes. While we find that coevolutionary patterns of the core circadian clock differ markedly in diapausing versus non-diapausing lineages, our analysis suggests that clock genes are just one component of a much broader network involving multiple interconnected genetic modules. This represents a major conceptual shift in the study of diapause genetics, moving beyond contemplating a "process governed by a small toolkit of timing genes" towards a polygenic and modular system that integrates metabolic, developmental, and regulatory networks arising from deep evolutionary history.

The macro-to-micro comparative framework developed here offers a novel route to establishing genotype-to-phenotype associations for complex traits that can be readily extended to other complex phenotypes. By applying this methodology to the diapause syndrome, which overlaps with other correlated traits, including adaptations to latitude, wet-dry seasonal cycles, migration, and life-history timing, we showcase the robustness and versatility of our framework in capturing both long-term evolutionary patterns and recent population-level variation. As high-quality genomic data continue to expand across the tree of life, future work incorporating denser genomic sampling within clades and broader ecological representation, coupled with similar comparative approaches, holds great promise for mechanistic dissection of complex phenotypes in a statistically rigorous way, even in non-model systems.

References

1. D. L. Denlinger, *Insect Diapause* (Cambridge university press, Cambridge, United Kingdom, 2022).

2. D. Posledovich, T. Toftegaard, C. Wiklund, J. Ehrlén, K. Gotthard, Latitudinal variation in diapause duration and post-winter development in two pierid butterflies in relation to phenological specialization. *Oecologia* **177**, 181–190 (2015).

3. V. Koštál, Eco-physiological phases of insect diapause. *Journal of Insect Physiology* **52**, 113–127 (2006).

4. M. J. Tauber, C. A. Tauber, S. Masaki, *Seasonal Adaptations of Insects* (Oxford Univ. Press, New York, 1986).

5. K. Wilsterman, M. A. Ballinger, C. M. Williams, A unifying, eco-physiological framework for animal dormancy. *Functional Ecology* **35**, 11–31 (2021).

6. G. J. Ragland, P. A. Armbruster, M. E. Meuti, Evolutionary and functional genetics of insect diapause: a call for greater integration. *Current Opinion in Insect Science* **36**, 74–81 (2019).

7. D. L. Denlinger, Insect diapause: from a rich history to an exciting future. *Journal of Experimental Biology* **226**, jeb245329 (2023).

8. K. J. Emerson, W. E. Bradshaw, C. M. Holzapfel, Complications of complexity: integrating environmental, genetic and hormonal control of insect diapause. *Trends in Genetics* **25**, 217–225 (2009).

9. D. Hernández-Carrasco, J. M. Tylianakis, D. A. Lytle, J. D. Tonkin, Ecological and evolutionary consequences of changing seasonality. *Science* **388**, eads4880 (2025).

10. E. J. Dowle, T. H. Q. Powell, M. M. Doellman, P. J. Meyers, M. B. Calvert, K. K. O. Walden, H. M. Robertson, S. H. Berlocher, J. L. Feder, D. A. Hahn, G. J. Ragland, Genome-wide variation and transcriptional changes in diverse developmental processes underlie the rapid evolution of seasonal adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23960–23969 (2020).

11. J. Rolland, L. F. Henao-Diaz, M. Doebeli, R. Germain, L. J. Harmon, L. L. Knowles, L. H. Liow, J. E. Mank, A. Machac, S. P. Otto, M. Pennell, N. Salamin, D. Silvestro, M. Sugawara, J. Uyeda, C. E. Wagner, D. Schluter, Conceptual and empirical bridges between micro- and macroevolution. *Nat Ecol Evol* **7**, 1181–1193 (2023).

12. A. Wagner, The molecular origins of evolutionary innovations. *Trends in Genetics* **27**, 397–410 (2011).

13. G. M. Kozak, C. B. Wadsworth, S. C. Kahne, S. M. Bogdanowicz, R. G. Harrison, B. S. Coates, E. B. Dopman, Genomic Basis of Circannual Rhythm in the European Corn Borer Moth. *Current Biology* **29**, 3501-3509.e5 (2019).

14. S. Zheng, Y. Wang, G. Li, S. Qin, Z. Dong, X. Yang, X. Xu, G. Fang, M. Li, S. Zhan, Functional polymorphism of CYCLE underlies the diapause variation in moths. *Science* **388**, eado2129 (2025).

15. P. Pruisscher, P. Lehmann, S. Nylin, K. Gotthard, C. W. Wheat, Extensive transcriptomic profiling of pupal diapause in a butterfly reveals a dynamic phenotype. *Molecular Ecology* **31**, 1269–1280 (2022).

16. M. J. Christmas, I. M. Kaplow, D. P. Genereux, M. X. Dong, G. M. Hughes, X. Li, P. F. Sullivan, A. G. Hindle, G. Andrews, J. C. Armstrong, M. Bianchi, A. M. Breit, M. Diekhans, C. Fanter, N. M. Foley, D. B. Goodman, L. Goodman, K. C. Keough, B. Kirilenko, A. Kowalczyk, C. Lawless, A. L. Lind, J. R. S. Meadows, L. R. Moreira, R. W. Redlich, L. Ryan, R. Swofford, A. Valenzuela, F. Wagner, O. Wallerman, A. R. Brown, J. Damas, K. Fan, J. Gatesy, J. Grimshaw, J. Johnson, S. V. Kozyrev, A. J. Lawler, V. D. Marinescu, K. M. Morrill, A. Osmanski, N. S. Paulat, B. N. Phan, S. K. Reilly, D. E. Schäffer, C. Steiner, M. A. Supple, A. P. Wilder, M. E. Wirthlin, J. R. Xue, Zoonomia Consortium§, B. W. Birren, S. Gazal, R. M. Hubley, K.-P. Koepfli, T. Marques-Bonet, W. K. Meyer, M. Nweeia, P. C. Sabeti, B. Shapiro, A. F. A. Smit, M. S. Springer, E. C. Teeling, Z. Weng, M. Hiller, D. L. Levesque, H. A. Lewin, W. J. Murphy, A. Navarro, B. Paten, K. S. Pollard, D. A. Ray, I. Ruf, O. A. Ryder, A. R. Pfenning, K. Lindblad-Toh, E. K. Karlsson, G. Andrews, J. C. Armstrong, M. Bianchi, B. W. Birren, K. R. Bredemeyer, A. M. Breit, M. J. Christmas, H. Clawson, J. Damas, F. Di Palma, M. Diekhans, M. X. Dong, E. Eizirik, K. Fan, C. Fanter, N. M. Foley, K. Forsberg-Nilsson, C. J. Garcia, J. Gatesy, S. Gazal, D. P. Genereux, L. Goodman, J. Grimshaw, M. K. Halsey, A. J. Harris, G. Hickey, M. Hiller, A. G. Hindle, R. M. Hubley, G. M. Hughes, J. Johnson, D. Juan, I. M. Kaplow, E. K. Karlsson, K. C. Keough, B. Kirilenko, K.-P. Koepfli, J. M. Korstian, A. Kowalczyk, S. V. Kozyrev, A. J. Lawler, C. Lawless, T. Lehmann, D. L. Levesque, H. A. Lewin, X. Li, A. Lind, K. Lindblad-Toh, A. Mackay-Smith, V. D. Marinescu, T. Marques-Bonet, V. C. Mason, J. R. S. Meadows, W. K. Meyer, J. E. Moore, L. R. Moreira, D. D. Moreno-Santillan, K. M. Morrill, G. Muntané, W. J. Murphy, A. Navarro, M. Nweeia, S. Ortmann, A. Osmanski, B. Paten, N. S. Paulat, A. R. Pfenning, B. N. Phan, K. S. Pollard, H. E. Pratt, D. A. Ray, S. K. Reilly, J. R. Rosen, I. Ruf, L. Ryan, O. A. Ryder, P. C. Sabeti, D. E. Schäffer, A. Serres, B. Shapiro, A. F. A. Smit, M. Springer, C. Srinivasan, C. Steiner, J. M. Storer, K. A. M. Sullivan, P. F. Sullivan, E. Sundström, M. A. Supple, R. Swofford, J.-E. Talbot, E. Teeling, J. Turner-Maier, A. Valenzuela, F. Wagner, O. Wallerman, C. Wang, J. Wang, Z. Weng, A. P. Wilder, M. E. Wirthlin, J. R. Xue, X. Zhang, Evolutionary constraint and innovation across hundreds of placental mammals. *Science* **380**, eabn3943 (2023).

17. S. Halali, E. Yapar, C. W. Wheat, N. Wahlberg, K. Gotthard, N. Chazot, S. Nylin, P. Lehmann, Tempo and mode of winter diapause evolution in butterflies. *Evolution Letters* **9**, 125–136 (2025).

18. C. J. Wright, L. Stevens, A. Mackintosh, M. Lawniczak, M. Blaxter, Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nat Ecol Evol* **8**, 777–790 (2024).

19. The Darwin Tree of Life Project Consortium, M. Blaxter, N. Mieszkowska, F. Di Palma, P. Holland, R. Durbin, T. Richards, M. Berriman, P. Kersey, P. Hollingsworth, W. Wilson, A. Twyford, E. Gaya, M. Lawniczak, O. Lewis, G. Broad, K. Howe, M. Hart, P. Flicek, I. Barnes, Sequence locally, think globally: The Darwin Tree of Life Project. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2115642118 (2022).

20. C. Wright, N. Wahlberg, R. Vila, M. Mutanen, P. Matos-Maraví, K. Lucek, I. Kleckova, L. Dapporto, V. Dinca, C. Bruschini, C. Wheat, M. Vila, L. Torrado-Blanco, V. Todisco, M. Rindos, P. Nguyen, P. Mulhair, S. Kamenova, M. Hicks, M. Espeland, I. Drinnenberg, M. Doblas-Bajo, R. Bailey, P. P. Community, M. Blaxter, J. Meier, Project Psyche: Generating and utilising reference genomes for all Lepidoptera in Europe. Biodiversity [Preprint] (2025). https://doi.org/10.32942/X2S35P.

21. C. Bortoluzzi, C. J. Wright, S. Lee, T. Cousins, T. A. L. Genez, D. Thybert, F. J. Martin, L. Haggerty, The Darwin Tree of Life Project Consortium, M. Blaxter, R. Durbin, Lepidoptera genomics based on 88 chromosomal reference sequences informs population genetic parameters for conservation. [Preprint] (2023). https://doi.org/10.1101/2023.04.14.536868.

22. A. Kowalczyk, W. K. Meyer, R. Partha, W. Mao, N. L. Clark, M. Chikina, RERconverge: an R package for associating evolutionary rates with convergent traits. *Bioinformatics* **35**, 4815–4817 (2019).

23. A. Kowalczyk, R. Partha, N. L. Clark, M. Chikina, Pan-mammalian analysis of molecular constraints underlying extended lifespan. *eLife* **9**, e51089 (2020).

24. E. Dalla Benetta, L. W. Beukeboom, L. Van De Zande, Adaptive Differences in Circadian Clock Gene Expression Patterns and Photoperiodic Diapause Induction in *Nasonia vitripennis*. *The American Naturalist* **193**, 881–896 (2019).

25. M. Jin, Y. Peng, J. Peng, S. Yu, C. Wu, X. Yang, J. Zhu, O. Infante, Q. Xu, H. Wang, K. Wu, Y. Xiao, A supergene controls facultative diapause in the crop pest Helicoverpa armigera. *Cell Reports* **43**, 114939 (2024).

26. N. L. Clark, E. Alani, C. F. Aquadro, Evolutionary rate covariation reveals shared functionality and coexpression of genes. *Genome Res.* **22**, 714–720 (2012).

27. J. Little, M. Chikina, N. L. Clark, Evolutionary rate covariation is a reliable predictor of co-functional interactions but not necessarily physical interactions. *eLife* **12**, RP93333 (2024).

28. D. Brady, A. Saviane, S. Cappellozza, F. Sandrelli, The Circadian Clock in Lepidoptera. *Front. Physiol.* **12**, 776826 (2021).

29. J. Kotwica-Rolinska, L. Chodáková, V. Smýkal, M. Damulewicz, J. Provazník, B. C.-H. Wu, M. Hejníková, D. Chvalová, D. Doležel, Loss of Timeless Underlies an Evolutionary Transition within the Circadian Clock. *Molecular Biology and Evolution* **39**, msab346 (2022).

30. K. Gotthard, C. W. Wheat, Diapause: Circadian Clock Genes Are at It Again. *Current Biology* **29**, R1245–R1246 (2019).

31. I. M. Aalberg Haugen, K. Gotthard, Diapause induction and relaxed selection on alternative developmental pathways in a butterfly. *Journal of Animal Ecology* **84**, 464–472 (2015).

32. O. Lindestad, C. W. Wheat, S. Nylin, K. Gotthard, Local adaptation of photoperiodic plasticity maintains life cycle variation within latitudes in a butterfly. *Ecology* **100**, e02550 (2019).

33. I. Siemers, L. Von Schmalensee, S. Ziesemer, K. Gotthard, P. Lehmann, Predictable local adaptation in butterfly photoperiodism but not thermal performance along a latitudinal cline. *Ecosphere* **15**, e4842 (2024).

34. M. Ittonen, A. Hagelin, C. Wiklund, K. Gotthard, Local adaptation to seasonal cues at the fronts of two parallel, climate-induced butterfly range expansions. *Ecology Letters* **25**, 2022–2033 (2022).

35. W. E. Bradshaw, M. C. Fletcher, C. M. Holzapfel, Clock-talk: have we forgotten about geographic variation? *J Comp Physiol A* **210**, 649–666 (2024).

36. S. G. Goto, Photoperiodic time measurement, photoreception, and circadian clocks in insect photoperiodism. *Appl Entomol Zool* **57**, 193–212 (2022).

37. D. S. Saunders, Dormancy, Diapause, and the Role of the Circadian System in Insect Photoperiodism. *Annu. Rev. Entomol.* **65**, 373–389 (2020).

Acknowledgments:

We thank Alexandra Hagelin, Matthew Nielsen, and Christer Wiklund for help with catching wild butterflies and Carolin Gierer for help with preparing samples for sequencing. We thank Project Psyche (https://www.projectpsyche.org/) for sharing their genomic resources for this work. We acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and NAISS for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. No AI-assistance was used in the writing of this manuscript.

Funding:

Carl Tryggers Foundation 17:163 (KG), Wenner-Gren Foundations UPD2023-0064 (CW), Swedish Research Council 2022-04507 (CW) and 2021-04258 (KG), Bolin Centre for Climate Research (KG, MI), Marie Skłodowska-Curie Actions Postdoctoral fellowship (Grant Agreement ID: 101104682) (SH)

Author contributions:

Conceptualization: SB, CW

Methodology: SB, CW

Investigation: SBB, DLA, MPW, WCB

Visualization: SB, SH

Sample acquisition: MI, KF, KG

Funding acquisition: CW, KG, MI

Supervision: CW

Writing – original draft: SB, CW

Writing – review & editing: SB, CW, SH, MI, KG

Competing interests: Authors declare that they have no competing interests.

Data and materials availability:

Sequence read data are available from NIH Short Read Archive under the following accessions: *Lasiommata mergera* (PRJNA1371628), *Pararge aegeria* (PRJEB49416 and PRJNA484116), *Pieris napi* (PRJNA449143), and *Papilio polytes (*PRJNA1166847). Scripts and intermediate files are deposited in Zenodo for reproducibility (10.5281/zenodo.17776754).

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S15

Tables S1 to S16

References

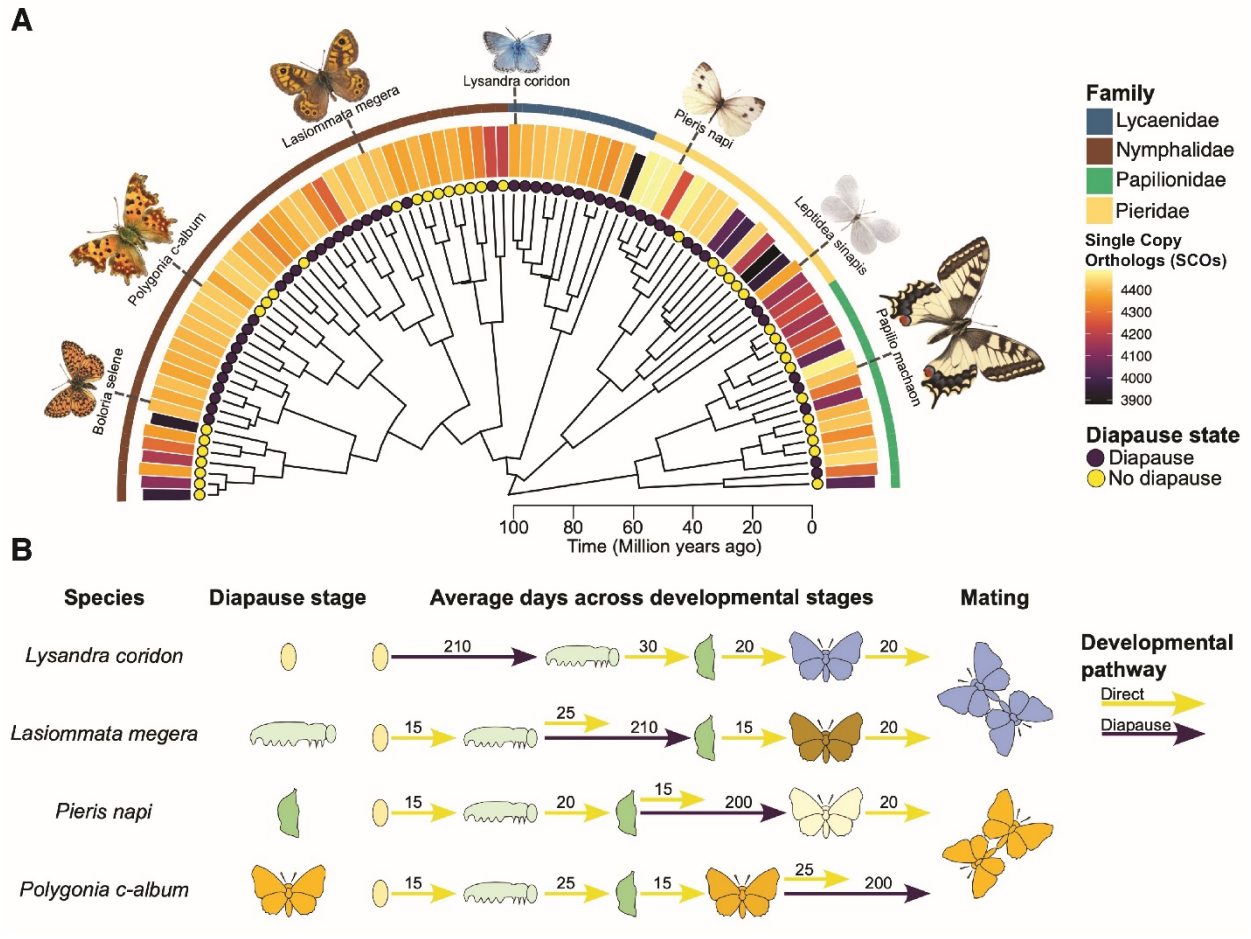

Fig. 1. Convergent evolution of diapause across butterflies at multiple developmental stages. (**A**) The evolutionary relationships among studied butterfly species in a time-calibrated, radial phylogeny. Concentric tracts display, from the tip (1) the diapause phenotype (purple - diapause and yellow – no diapause), (2) the number of single copy orthologous genes used per species, and (3) the butterfly family classification. (**B**) Comparison of average lifespans in exemplar butterfly species in our dataset that exhibit diapause at different developmental stages, compared with direct development when species have facultative diapause induction (i.e., both pathways).

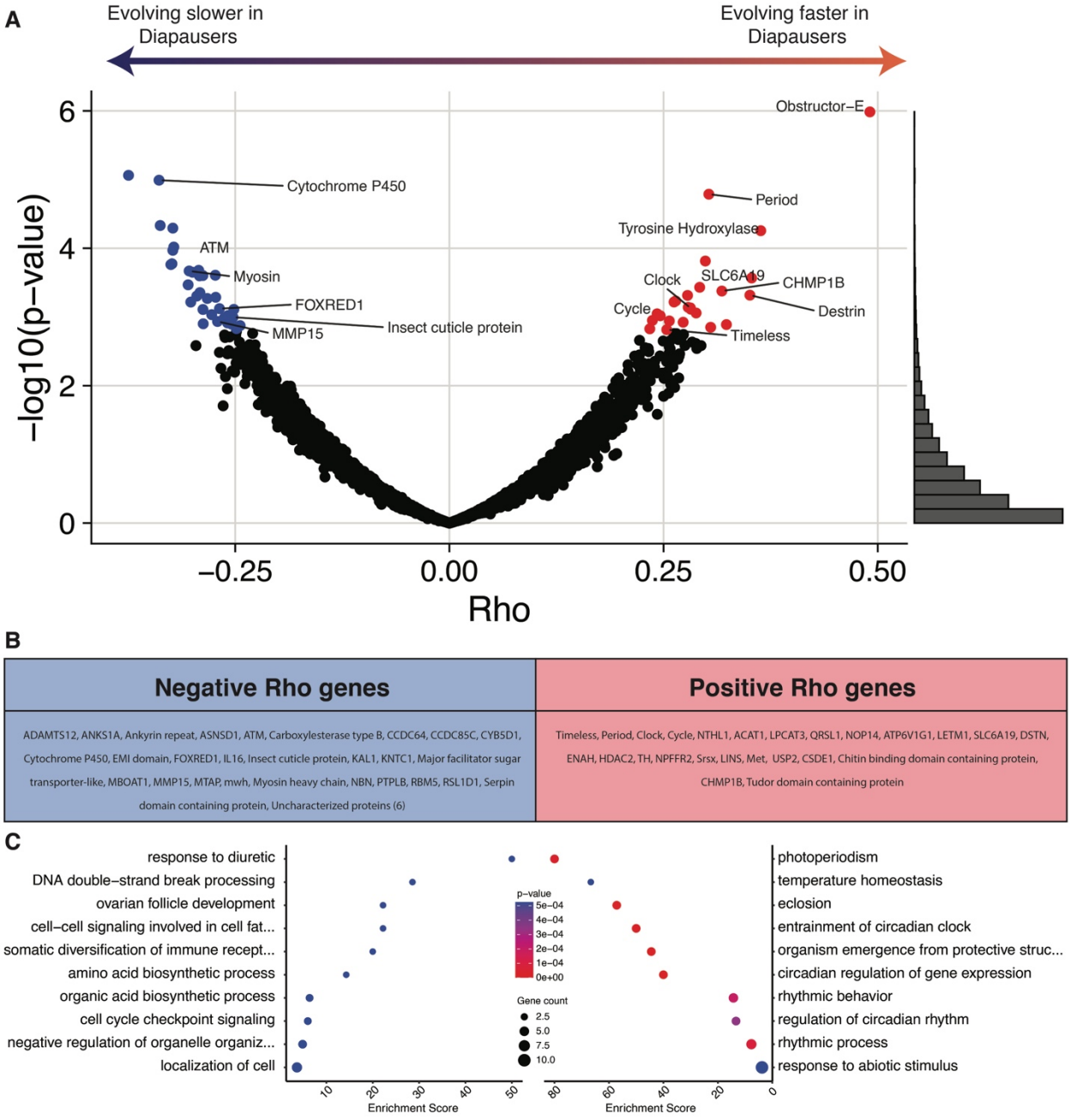

Fig 2. Correlation between relative evolutionary rates and diapause identifies candidate genes associated with diapause evolution across butterflies. (**A**) Distribution of correlation between relative evolutionary rates with diapause phenotype (Rho) across single copy genes. The histogram in the margin shows the spread of Rho values. Genes significantly correlated with diapause (q-value < 0.1; table S7), are highlighted (red: positive and blue: negative). (**B**) A list of genes with significant positive (red) and negative (blue) Rho values. (**C**) Gene Ontology enrichment analysis of the positively and negatively correlated gene sets. Dot plots display enriched biological processes, colored by p-value and sized by gene count. Positively associated genes are enriched for circadian processes (right), while negatively associated genes are linked to metabolism and development (left).

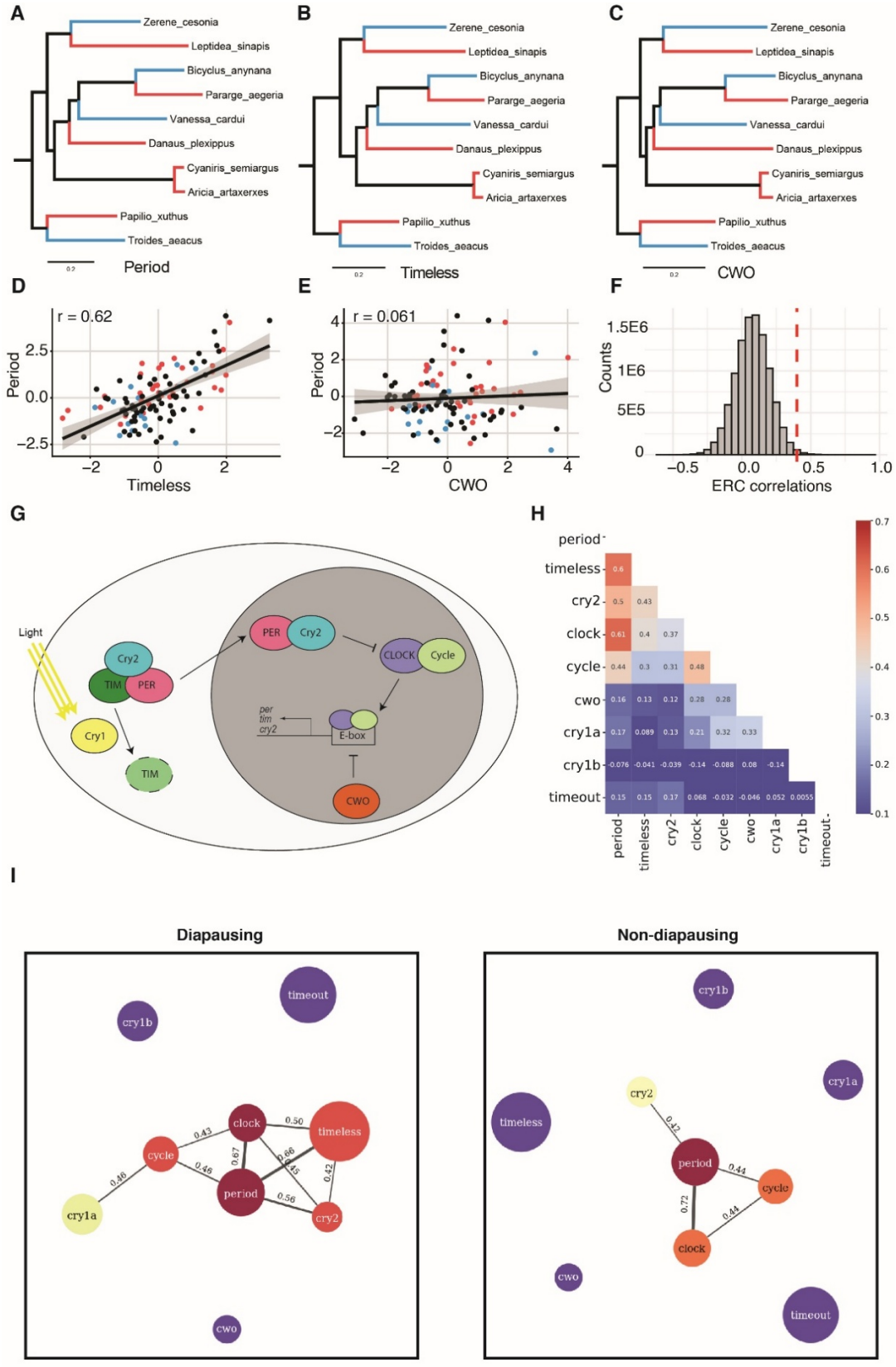
A
Zerene_cesonia
Leptidea_sinapis
Bicyclus_anynana
Pararge_aegeria
Vanessa_cardui
Danaus_plexippus
Cyaniris_semiargus
Aricia_artaxerxes
Papilio_xuthus
Troides_aeacus
Period
B
Timeless
C
CWO
D
r = 0.62
Period
Timeless
E
r = 0.061
CWO
F
Counts
1.5E6
1E6
5E5
0
ERC correlations
G
Light
Cry2
TIM
PER
Cry1
PER
Cry2
CLOCK
Cycle
per
tim
cry2
E-box
CWO
H
period
timeless
cry2
clock
cycle
cwo
cry1a
cry1b
timeout
I
Diapausing
Non-diapausing

Fig 3. Evolutionary dynamics of circadian clock components differ between diapausing and non-diapausing butterflies. (**A-C**) Evolution rates on a subset of species trees for *period*, *timeless*, and *cwo*, colored by diapause phenotype (red = diapausing, blue = non-diapausing). (**D-E**) Pairwise rate correlations per branch between core circadian genes (*period* vs. *timeless*, and *period* vs. *cwo*), documenting differences in rate covariation (shown without phylogenetic correction). (**F**) Distribution of ERC values across all gene pairs (n=10,444,735), highlighting the abundance of low correlation values and our coevolving cut-off threshold (ERC = 0.4, top 0.52 % of interactions). (**G**) Schematic of the Lepidopteran circadian clock network, showing known molecular interactions between clock proteins and their feedback loops(*28*). (**H**) Pairwise ERC matrix among circadian clock genes, highlighting variation in their co-evolutionary interactions. (**I**) Network diagrams of circadian clock genes (nodes: circles), where genes are connected when ERC > 0.4 (edges with ERC value), for the diapausing (left) and non-diapausing (right) datasets. Diapausing species exhibit a larger network of ERC connectivity among core circadian clock genes, suggesting strong coevolution under selective pressure related to the diapause syndrome.

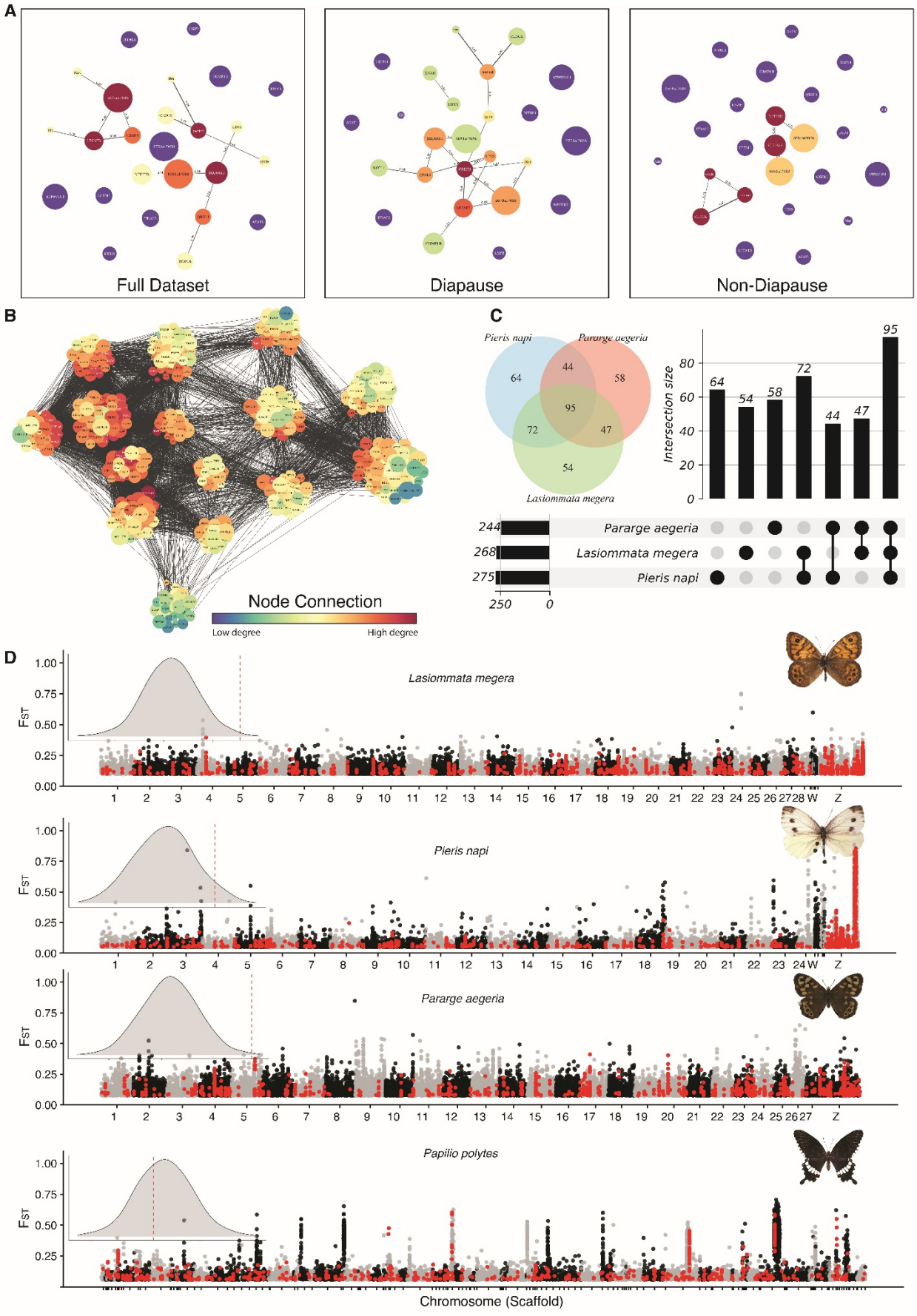
A
Full Dataset
Diapause
Non-Diapause
B
Node Connection
Low degree
High degree
C
Pieris napi
Pararge aegeria
Lasiommata megera
64
44
58
95
72
47
54
Intersection size
95
64
54
58
72
44
47
244
268
275
250
0
D
Lasiommata megera
Pieris napi
Pararge aegeria
Papilio polytes
Fst
Chromosome (Scaffold)

Fig 4. Co-evolutionary network of diapause-associated genes and cross-species comparison of their variation in butterflies

(**A**) ERC networks among genes significantly associated with diapause (n = 25; Rho-genes with ERC > 0.4), constructed separately for diapausing, non-diapausing, and all species. Diapausing species exhibit markedly higher ERC connectivity among Rho-genes, forming the largest connected component comprising 14 genes. (**B**) ERC network of genes (n = 504; expanded Rho-gene network) co-evolving (ERC > 0.4) with those in the largest connected Rho-gene module (n = 14) from diapausing species, clustered using a stochastic block model (SBM). Node color represents connectivity (degree), from low (blue) to high (red), highlighting potential hub genes that may play key roles in diapause regulation. (**C**) UpSet plot and Venn diagram illustrating the number of overlapping genes among the species-specific intersection set (expanded Rho-gene network among outliers) across the three diapausing butterfly species, revealing that a substantial fraction of outlier genes are shared across three species with a diapause syndrome. (**D**) The top 5% F_{ST} outliers between two geographically distinct populations very likely to differ in critical photoperiod and diapause syndrome, across three butterfly species: *Lasiommata megera* (F_{ST} > 0.101), *Pieris napi* (F_{ST} > 0.05), *Pararge aegeria* (F_{ST} > 0.076), and two populations of *Papilio polytes* (F_{ST} > 0.058) which lacks a diapause syndrome. Genes from the expanded Rho-gene network are highlighted in red indicating their location among outlier loci. For each species, an inset histogram shows the distribution of the number of genes in random sets (n = 504; 1000 replicates) drawn from the SCOs dataset that intersect with outlier loci. The dotted red line marks the position of the species-specific intersection set (expanded Rho-gene network among outliers), containing 268, 275, 244, and 303 genes corresponding to the p-values 0.006, 0.056, 0.002, and 0.66 for *L. megera*, *P. napi*, *P. aegeria*, and *P. polytes*, respectively.