Project Psyche: Generating and utilising reference genomes for all Lepidoptera in Europe

Charlotte J. Wright*[1], Niklas Wahlberg[2], Roger Vila[3], Marko Mutanen[4], Pável Matos-Maraví[5], Kay Lucek[6], Irena Kleckova[5], Leonardo Dapporto[7], Vlad Dincă[8], Claudia Bruschini[7], Christopher W. Wheat[9], Marta Vila[10], Laura Torrado-Blanco[10,11], Valentina Todisco[12], Michal Rindos[5], Petr Nguyen[5,13], Peter O. Mulhair[14], Stefaniya Kamenova[15], Marcus Hicks[16], Marianne Espeland[17], Ines A. Drinnenberg[18], Mónica Doblas-Bajo[11], Richard I. Bailey[19], The Project Psyche Community[20], Mark Blaxter[1], Joana I. Meier*[1,21]

Affiliations:

[1] Tree of Life Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom

[2] Department of Biology, Lund University, 223 62 Lund, Sweden

[3] Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Martítim de la Barceloneta 37, 08003 Barcelona, Spain

[4] Ecology and Genetics Research Unit, University of Oulu, Finland, Oulu, Finland

[5] Biology Centre CAS, Institute of Entomology, Branišovská 1160/31, 370 05 České Budějovice, Czech Republic

[6] Biodiversity Genomics Laboratory, Institute of Biology, University of Neuchâtel, 2000, Neuchâtel

[7] Biology Department, Università degli Studi di Firenze, via Madonna del Piano 6, 50019, Sesto Fiorentino, Italy

[8] "Grigore Antipa" National Museum of Natural History, Bucharest, Romania

[9] Department of Zoology, Stockholm University, Stockholm, Sweden

[10] Departamento de Bioloxía, Facultade de Ciencias, Universidade da Coruña, Campus da Zapateira, 15071 A Coruña, Spain

[11] Zoology Unit, Department of Organisms and Systems Biology, University of Oviedo, Campus del Cristo, C/Catedrático Valentín Andrés Álvarez, 33006 Asturias, Spain

[12] Department of Environment and Biodiversity, University of Salzburg, Hellbrunnerstraße 34, Salzburg, 5020, Austria

[13] University of South Bohemia, Faculty of Science, Branišovská 1760, 370 05 České Budějovice, Czech Republic

[14] Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, L59 7ZB, UK

[15] National Museum of Natural History, Bulgarian Academy of Sciences, 1 Tsar Osvoboditel Blvd., 1000 Sofia, Bulgaria

[16] School of Biological and Behavioural Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, United Kingdom

[17] Leibniz Institute for the Analysis of Biodiversity Change (LIB), Museum Koenig, Adenauerallee 127, 53113 Bonn, Germany

[18] Institut Curie, PSL University, Sorbonne Université, CNRS, Nuclear Dynamics, 75005 Paris, France

[19] Department of Ecology and Vertebrate Zoology, Faculty of Biology and Environmental Protection, University of Lodz, Łód´z, Poland

[20] The Project Psyche Community: https://zenodo.org/records/15359862

[21] Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, United Kingdom

Corresponding authors: charlotte.wright@sanger.ac.uk and joana.meier@sanger.ac.uk

Keywords: Lepidoptera, butterflies, moths, reference genomes, evolution, conservation

Abstract

Project Psyche is a trans-national initiative to generate and study chromosome-level reference genomes of all ca. 11,000 described species of Lepidoptera (butterflies and moths) found in Europe. The Project Psyche community encompasses diverse researchers, amateur lepidopterists, practitioners, and industry experts united by a common vision of the importance of genomics for Lepidoptera. Lepidoptera are at the forefront of biodiversity genomics with genomes for over 1,000 species already generated - the highest number across all eukaryotic orders. Here, we outline how Project Psyche is generating and harnessing these genomes. This pan-European catalogue of openly accessible lepidopteran genomes will transform our understanding of evolution and ecology, inform conservation, and foster advances in pest management.

Highlights

- Project Psyche aims to generate openly accessible, chromosome-level reference genomes for all 11,000 species of Lepidoptera occurring in Europe, with 1,000 genomes already completed.
- This ambitious goal will be achieved by a decentralised network of collection and sequencing hubs across Europe. Genome analysis will be conducted in a highly collaborative and inclusive way.
- Thousands of lepidopteran genomes will propel diverse research areas including comparative genomics, phylogenomics, molecular evolution and population genomics.
- Lepidoptera are key indicators of ecosystem health. Genomes will facilitate monitoring and the conservation of Lepidoptera, which in turn will protect ecosystems.
- Lepidoptera include many species of economic and agricultural importance. The genomes will fuel the development of more effective pest management strategies.

Introduction

With approximately 160,000 described species [1], Lepidoptera encompass ~10% of known species. This mega-diversity is the product of over 230 million years of evolution [2–4], with numerous species central to many vital ecological functions, including herbivory and pollination. Lepidoptera comprise flagship species for conservation and the protection of entire ecosystems [5]. Some species represent devastating agricultural, forestry and textile pests [6,7]. Their spectacular phenotypic diversity, intimate interactions with host plants and social insects, and cultural importance have fascinated researchers, hobbyists and the public alike for centuries, making them ideal for citizen science and connecting society with nature. This long-term interest in butterflies and moths has made them one of the best-studied orders, with detailed knowledge of their taxonomy, life history, distribution, demography, and ecology. The lepidopteran research community has been proactive in adopting genomics to

understand the evolution and ecology of specific species, including the genetics of wing patterns [8], the impacts of climate change [9], and the biology of migration [10]. Coupled with citizen recordings and extensive long-term monitoring programs, genomes allow us to gain unique insights into the drivers of ever faster insect declines and ecosystem changes [11]. Moreover, many natural history collections, which encompass millions of Lepidoptera specimens, are now being catalogued [12], offering opportunities for integrating insights from museum collections with genomic studies.

Reference genomes are the foundation for all genomic studies. They unite **microevolutionary** studies using short-read resequencing, RNA or epigenetics data that are mapped to the reference genome, while providing connections to **macroevolutionary** studies via phylogenomics and comparative genomics. Generating chromosomally-resolved, high-quality genomes from the outset, rather than producing a larger number of cheaper, low-quality genomes, provides broad, future-proof utility. Due to ongoing technological advances and reduced sequencing costs, generating thousands of chromosome-level genomes is now feasible. Indeed, the Earth Biogenome Project (EBP) aims to generate reference genomes of all eukaryotic life on Earth [13]. This moonshot goal will be reached through the collaborative efforts of many biodiversity genomics initiatives. Here we present Project Psyche, an ambitious trans-national project to generate **chromosome-level reference genomes** for all ca. 11,000 species of Lepidoptera occurring in Europe. The project was named after the Greek goddess of the soul, Psyche, depicted with butterfly wings. Project Psyche strives to make all reference genomes and analytic datasets openly available and easily accessible to all. Through a decentralised workflow for genome generation and analysis, together with early-career researcher training, Project Psyche promotes equity in biodiversity genomics.

Lepidopteran genomes are often small (~500 Mb) and their **holocentric chromosomes** [14] lack large, often difficult-to-assemble localised centromeric regions [15]. These features permit a standardised, large-scale approach to genome assembly and analysis, enabling cross-species comparisons with minimised methodological biases. Project Psyche also benefits from the progress made by other large-scale biodiversity sequencing projects such as the Darwin Tree of Life (DToL) project [16] and the European Reference Genome Atlas (ERGA) [17], that have developed standardised, high-throughput methods for sequencing biodiversity.

Generating chromosome-level reference genomes across an entire, highly diverse order represents a paradigm change for biology. Reference genomes for all species of Lepidoptera in Europe will enable the exploration of general principles in evolutionary biology and ecology, and allow pressing societal challenges to be addressed. Here, we highlight how Psyche genomes will deepen our understanding of the drivers of species and ecological diversification, genome and co-biont evolution, and the evolutionary basis of adaptation. We also discuss the potential of genomics for conservation and new solutions to societal problems. Accelerating research in these areas is timely given the pressing threats of global change and biodiversity loss.

The importance of natural history and taxonomy

Sequencing high-quality genomes of all species of a diverse order across an entire continent is a formidable endeavour involving collective efforts and careful planning **(Box 1)**. Currently 11,665 species of Lepidoptera have been recorded in Europe [18], although their systematics is a dynamic field of research, with revisions and descriptions of new species being regularly published. Project Psyche fosters reciprocal knowledge exchange between systematists and genomicists, as generating reference genomes relies on taxonomic expertise to sample and identify specimens. Conversely, reference genomes, combined with morphological and biogeographical data, help to resolve taxonomic uncertainties.

Due to the decline of taxonomists [19], the successful sampling of all Lepidoptera species in Europe requires close engagement with amateur lepidopterists and their detailed knowledge of local faunas, knowledge that is rarely documented in a formal way. Close collaboration between taxonomists and genomicists can identify essential knowledge about phenotypic and behavioural variation, which opens novel research directions. This includes resolving taxonomic uncertainties, understanding host specialisation and elucidating genotype-phenotype associations. Engagement with natural history museums and their curators [20,21] provides further historical and contemporary data to complement field observations and genomic data. Project Psyche strives to be open and inclusive, encouraging new members, particularly from under-represented countries and regions **(Box 1)**. While sequencing is Europe-focused, the genome users are global.

The rank abundance and distribution of Lepidoptera in Europe forms a continuum from common species that are easy to identify and collect, to rare or localised species often belonging to difficult species complexes, requiring specialised taxonomic knowledge and/or barcoding for identification. Among the first 1,000 genomes, Project Psyche has covered many of the easily identifiable and widespread taxa. Joint collecting trips will target regions with high endemicity, such as the Balkans, the Iberian Peninsula, the Alps, and the Mediterranean islands. Finally, challenging species complexes will be resolved in close partnership with taxonomists, and complemented by population genomic studies to inform species' delimitation, and inform an updated European checklist.

Sequencing Lepidoptera at scale

Project Psyche is committed to ethical and sustainable sampling practices [17]. We use a standardised methodology from sample collection to genome generation (DOI: dx.doi.org/10.17504/protocols.io.yxmvmbk1bg3p/v1) **(Figure 1)**, building on protocols developed by the DToL project [22]. In brief, sampling aims for 1-5 females depending on body size. Females are preferred because they are the **heterogametic sex** in Lepidoptera, and thus allow the identification and assembly of all sex chromosomes. Collected specimens are dissected on dry ice and snap-frozen in liquid nitrogen or -80°C freezers. Crucially, specimens remain deep-frozen until DNA extraction. The continuity of the cold chain is essential, although tests of alternative DNA preservation methods are underway **(Box 2)**. Typically, the head is used for **long-range sequencing** (currently Hi-C), the thorax for **long-read sequencing** (currently Pacific Biosciences, PacBio) and the abdomen or another individual for gene annotation using RNA sequencing [23] **(Figure 1A)**. For some

specimens, genitalia, or legs for **DNA barcoding**, may also be removed to confirm identification. For small specimens, the entire body may be needed for a single type of sequencing; hence several specimens are required. Wings provide voucher material that will be stored at local natural history museums and their photos are made available on BioImage Archive (https://doi.org/10.6019/S-BIAD1504) [24]. Databases and online systems are used to track sampling status with standardised metadata in real time. To facilitate coordinated sample collections both within Psyche and with other biodiversity genomics projects, we have developed a portal (https://psyche.tol.sanger.ac.uk/) that shows the status for each species. Our species list is also integrated into the Genomes on a Tree (GoaT) database (https://goat.genomehubs.org) [25], which is used to coordinate ongoing reference genome sequencing efforts across many projects, to minimise duplication of genome sequencing. GoaT also provides an estimate of the genome size, and thus the amount of sequencing data needed, and the expected chromosome number for a given species. Currently, PacBio sequencing is done to 25-30-fold coverage (>20 kb fragment lengths) and we generate 100 Gbp Illumina data for Hi-C.

We use quality assessments throughout the production pipeline to ensure adherence to quality standards. Symbionts, parasites and other microbiota are frequently sequenced alongside Lepidoptera [26,27]. Their genomes are assembled separately from the lepidopteran mitochondrial and nuclear genome. Next, the nuclear genome is assembled and **scaffolded** from the long-read and long-range (Hi-C) data **(Figure 1B)**, using EBP best-practice approaches. The high heterozygosity of lepidopteran genomes [28] allows using new algorithms to separate the data into two **haplotypes** (whenever PacBio and Hi-C data come from the same individual). These two haplotypes are manually curated, improving scaffolding accuracy to produce a haplotype-phased chromosome-level assembly that meets EBP standards ([29]; https://www.earthbiogenome.org/). This time-consuming manual task is achieved at scale through decentralised curation of the genomes by Project Psyche members across Europe, trained by the Wellcome Sanger Institute (WSI) curation team **(Box 2)**. Once complete, all components (both nuclear haplotypes, mitochondrial genome, cobiont genomes, raw data), are submitted to the European Nucleotide Archive (ENA) [30]. In addition, DNA barcode sequences used to confirm specimen identification are uploaded to the Barcode of Life Data Systems (BOLD) [31]. Each genome is announced to the community through a Genome Note, which describes the genome and credits sample collectors and genome curators through authorship (https://wellcomeopenresearch.org/gateways/treeoflife/projectpsyche). Standardised analytical resources are being generated for each genome and for sets of genomes **(Figure 1C)**, including gene annotations, genome metrics, k-mer profiles, sequence alignments, variant calls and repeat annotations. These resources are made publicly available and can be interactively examined and compared (https://gap.cog.sanger.ac.uk; www.lepbase.org). The development of these open resources is driven by the Project Psyche community to ensure the resources meet researcher needs while facilitating collaborative science.

Strengthening the genomics community across Europe and beyond

Project Psyche has made rapid progress since launching in Summer 2023, having achieved its first milestone of 1,000 assembled genomes and 2,000 species collected **(Figure 2, Supplementary Figure 1)**. Note, the number of genomes includes 512 species sequenced

at the WSI as part of the DToL project. Similarly, species sequenced by Project Psyche also contribute to the aims of other projects including ERGA and EBP. Currently, Lepidoptera are the most extensively sequenced animal order **(Figure 2A)**. Project Psyche is in its first of three phases. Phase 1 aims to generate genomes for 2,000 species of Lepidoptera in Europe. Specimens are collected through seven regional sample collection hubs across Europe, and sequencing is carried out at the WSI, which is developing high-throughput and low-input sequencing methods **(Figure 2C)**. We also develop protocols and approaches to support community growth and include more collectors and sequencing locations. In Phase 1, we strive to sample representatives of all 92 families and 27 superfamilies present in Europe. In Phase 2, we aim to sequence an additional 5,000 species, including whole ecosystem sampling for selected locations. In anticipation of this, Project Psyche has recently established sub-hubs that serve as local collection points that contribute to a regional collection hub (**Box 2**). For Phase 3, the ambition is to sequence most of the remaining ca. 4,000 species while further increasing community participation. Genomes of Lepidoptera from Europe generated by individual research groups can also be included under the Project Psyche umbrella provided they fulfil EBP standards ([29]; https://www.earthbiogenome.org/). Project Psyche aims to develop genomics capacity across Europe and beyond. This includes sharing of protocols and systems, training, mentorship and collaborative grants. We envision that Project Psyche will motivate and empower similar initiatives for other taxonomic groups and geographic regions.

The scientific potential of 11,000 lepidopteran genomes

Lepidoptera are a powerful model for understanding genome evolution, the genetic basis of phenotypic diversity, species richness, and the ecological dynamics of a mega-diverse clade. By providing large-scale, standardised genomic resources, Project Psyche allows for the extensive ecological, life history, phenological, and distributional understanding of Lepidoptera to be integrated for insightful evolutionary inquiry. Crucially, thousands of chromosome-level assemblies increase accuracy in detecting and determining genomic changes (e.g. in gene content, transposable elements, or chromosomal rearrangements) associated with phenotypic divergence over 230 million years of evolution. The assemblage of lepidopteran species found in Europe constitutes ~7% of all lepidopteran species globally. This non-random sampling of evolutionary diversity [1] highlights the need for similar regional initiatives. Nonetheless, with representatives from the 92/136 families and 27/46 superfamilies that occur in Europe [18], the Project Psyche dataset is an unprecedented resource for addressing evolutionary questions **(Figure 2B).**

Genome evolution

Large-scale genomic datasets allow the systematic study of chromosomal evolution across Lepidoptera. Lepidoptera are known to vary dramatically in rates of karyotypic evolution, which may have influenced diversification rates by generating reproductive isolation barriers. For instance, most Lepidoptera have 30-31 chromosomes [32,33], which closely resemble the inferred ancestral linkage groups known as Merian elements [34]. Most species have experienced few chromosome fusions, while fissions are restricted to some genera. However, several lineages have recently undergone extensive genomic rearrangements, with or without dramatic changes in chromosome number [34]. Thousands of reference genomes will enable detailed comparisons between close relatives that differ in their genome structure **(Figure 3A)**. This will shed light on the mechanistic basis and cause of chromosome rearrangements, the evolutionary constraints on rearrangement [35], and the effects of these chromosomal rearrangements on recombination, adaptation and speciation.

A key question is how holocentricity evolved in Lepidoptera and its sister order Trichoptera. Characterisation of the holocentromere in the silk moth*, Bombyx mori*, found that a key component of eukaryotic centromeres, CENP-A/CenH3, has been lost [36], while many others are conserved [37]. A diverse set of reference genomes will reveal the evolutionary dynamics of kinetochore composition, including the loss of CENP-A and possible loss of other centromeric components. The holocentric architecture requires meiotic adaptations such as restricted kinetochore activity during chromosome segregation [38,39], which affect the recombination landscape and, in turn, shape the distribution of genes and repeats across the genome. Thousands of genomes provide the foundations for characterising recombination landscapes across the order, revealing the genomic correlates of recombination rate variation.

W chromosomes are typically repeat-rich and historically were difficult to assemble. However, assembly of W chromosomes is now routine thanks to high-quality long reads, combined with long-range information and careful manual genome curation. The Project Psyche data will also allow us to resolve the longstanding debate of how W chromosomes in Lepidoptera arise [40]. Hypotheses include the emergence via the adoption of a **B chromosome**, acquisition of a novel female-determining locus by another chromosome, or

fusion of sex chromosomes with autosomes [41,42]. Current evidence suggests that W chromosomes evolved multiple times independently in Lepidoptera, and were lost in some species [41]. Indeed, sex chromosome-autosome fusions are the most frequent type of inter-chromosomal rearrangement in Lepidoptera [34] and the resulting neo-sex chromosomes can retain homology in regions derived from original autosomes [35,43,44]. As female Lepidoptera likely do not recombine [45], W chromosomes are expected to degenerate over time. The availability of neo-sex chromosomes of various ages offers new opportunities to investigate the temporal dynamics of W chromosome degeneration and the consequences of Z-linked inheritance. Finally, W chromosome sequences will shed light on their role in sex determination and the evolution of diverse primary sex-determining mechanisms in Lepidoptera. More than one sex determination mechanism is known in Lepidoptera [46], including a dominant W-linked female determiner in *B. mori* [47], sex determination based on the Z chromosome to autosome ratio in *Samia cynthia* [48], and zygosity-based sex determination in a hypervariable Z locus in *Bicyclus anynana* [49].

The chromosome conformation (Hi-C) data used to scaffold Psyche genomes also allows for comparative analyses of genome organisation to include **three-dimensional (3D) chromatin architecture**. The high conservation of linear genome organisation across many lepidopteran species raises the question whether this conservation also applies to the 3D chromatin structure. Conversely, in lineages that have undergone extensive genomic rearrangements, it remains unclear whether changes in 3D genome organisation are a consequence of, or a constraint upon, alterations in linear genome structure. The spatial genome organisation of Lepidoptera is remarkable: Firstly, **chromosome territoriality** is extremely pronounced with limited inter-chromosomal contacts **(Figure 3B)** [50,51]. Secondly, in addition to the active A and inactive B compartments described in various invertebrates and vertebrates [52], *B. mori* chromosomes also fold into a third compartment, S, with a unique contact pattern [50,51]. The Psyche dataset will allow the investigation of whether strong territoriality and the S compartment are general characteristics of Lepidoptera, and conservation of other levels of organisation including **topologically associating domains** (TADs) **(Figure 3C)**. Ultimately, this information will facilitate the identification of the cellular mechanisms that underlie their formation, their functional relevance, and whether these patterns are functionally linked to evolutionary patterns in the linear genome.

Drivers of diversification in Lepidoptera

An outstanding question is how Lepidoptera became extremely species-rich and phenotypically diverse. An exceptional burst of diversification occurred ~100-150 MYA, where nearly all superfamilies in Ditrysia evolved, comprising 99% of all extant Lepidoptera species [4,53]. This diversification roughly coincides with the origin of flowering plants [2,54], although considerable temporal uncertainty remains. A broader diversity of reference genomes will allow us to infer more accurately dated phylogenies, drawing upon both coding and non-coding regions, **(Figure 3D, 3E)**, accounting for topological variation due to incomplete lineage sorting or introgression.

Why some lineages diversified much faster and more extensively than others remains an open question [55]. Factors that may facilitate rapid speciation include high genetic variation [56,57], chromosomal rearrangements [58], introgression [59], expansion of transposable elements [60] and gene birth/death dynamics [61]. By comparing representative genomes

from all Lepidoptera lineages in Europe that cover the spectrum of diversification rates, we can disentangle factors unique to rapidly evolving taxa from slower clades. For instance, introgression provided key genetic variation in many cases of rapid diversification [62], but it is unknown if introgression is equally common among slowly speciating lineages. Similarly, chromosomal rearrangements are thought to contribute to speciation through hybrid dysfunction or recombination suppression [63–65], but this is yet to be systematically investigated. Direct quantification of chromosomal rearrangement rates and comparisons between conserved and rearranged genomes [66], will allow us to investigate whether all or specific types of rearrangements generally impact diversification rates.

Molecular drivers of adaptation

Project Psyche will also fuel the study of the genetic basis of diverse adaptations. The genomic basis of novel traits have been elucidated in select species. For example, gene duplication enabled Pieridae to overcome the glucosinolate chemical defenses of their Brassicales host plants [67,68], while transposable element insertions have been found causing wing colour polymorphism in a wood tiger moth [69] and the peppered moth [70], and alternative life-history strategies in a clouded yellow butterfly [71]. These genomic changes can arise *de novo*, or through the exchange of genetic material. For example, horizontal gene transfer led to venom production in asp caterpillars (Megalopygidae) [72], whereas introgression facilitated pyrethroid insecticide resistance through gene flow from resistant invading *Helicoverpa armigera* into the native pest species *Helicoverpa zea* in Brazil [73]**.** With a large and densely sampled dataset of lepidopteran genomes, researchers will be able to upscale these analyses and assess their generality **(Figure 3G)**.

Lepidoptera in Europe display a huge range of life histories and phenologies, including associations with ants in Lycaenidae, keratin feeding in Tineidae, and shifts between nocturnal and diurnal activity in many groups. Reference genomes will empower researchers to embark upon mechanistic investigations of numerous evolutionary novelties. Importantly, these studies will benefit from standardised genome annotations for all species, which will facilitate quantification of gene birth/death dynamics that minimise bias from heterogeneous annotations [74], and references for mapping diverse transcriptomic data. By examining the rate and mode of gene gain, duplication and loss **(Figure 3H)**, and their consequences for gene expression **(Figure 3I)**, as well as shifts in selective pressures and recombination acting on genes during these transition events, we can begin to disentangle the genetic basis of adaptation and its evolution. Thousands of genomes, spanning a large range of divergence times, also increases our power to test major theoretical predictions including the repeatability of evolution at the genomic and phenotypic level [75,76]. For example, whether higher divergence time among taxa experiencing similar selection, predicts a lower likelihood of parallelism at the genetic level [77].

We lack a general understanding of how gene regulatory networks are structured and evolve in Lepidoptera, beyond a small number of intensely studied gene regulatory networks, such as those involved in wing colour patterning [78] and Hox genes [79]. Therefore, the extent to which regulatory networks pose a constraint on the reorganisation of genetic material is unknown. This is pertinent given the extremely high levels of conservation of gene order [34,80] but low sequence similarity between non-coding regions [28]. Non-coding regions can drive evolutionary novelty. For example, a long non-coding RNA generates a miRNA that drives wing colour variation in butterflies [81–83]. Through denser sampling of species,

the evolutionary history of non-coding sequences can be traced at a finer scale, leading to analyses of strong conservation associated with selective constraint and opening the door to understanding their function. In addition, chromosome conformation data will enable the investigation of whether regulatory rewiring, through changes in 3D architecture, contributes to trait evolution [84–86] **(Figure 3C)**.

Evolution of cobionts

Genome sequencing of wild specimens often provides information about their associated microbiomes, endosymbionts, parasites, viruses, and other **cobionts**. Reference genomes will fuel knowledge on the spectrum of cobionts associated with Lepidoptera. *Wolbachia* and *Spiroplasma* are bacteria commonly found in Lepidoptera that can manipulate sex ratios and cause cytoplasmic incompatibilities in their hosts [26,87,88]. The Project Psyche dataset will enable patterns of cobiont presence and host switching events to be mapped across Lepidoptera **(Figure 3F)**, building on previous work that indicated limited co-evolution between cobionts such as *Wolbachia* and their insect hosts [26]. This dataset will also enable investigations into the complex interplay of viral and host evolution. For example, *Wolbachia* can itself be parasitised by double-stranded DNA temperate bacteriophages [89].

Applications to conservation and society

There is an urgent need to understand and predict how, and how quickly, populations will adapt to a changing environment, allowing the persistence of ecological communities [90]. This is an issue more tractable in Lepidoptera given detailed information on current and past distributions, as well as biotic interactions and life histories. Additionally, Lepidoptera are often used as important indicators of ecosystem health [5]. Genomics can provide insight into the genomic bases of evolutionary responses to environmental change that may increase ecological resilience [91]. Psyche genomes also provide a foundation for understanding genetic diversity over time **(Figure 3J)**. A powerful approach will be to combine Psyche genomes with re-sequencing data from many individuals to trace genetic diversity across regions and time, via museum collections **(Figure 3K, Figure 3L)**. This approach will be employed by projects including the European Lepidopteran Population Genomics Consortium (LepEU; https://lepeu.github.io/). This will provide insights into population sizes, connectivity, and local adaptation and inform conservation management **(Figure 3P)** [92]. There is growing consideration of population structure and genetic diversity in addition to taxonomic units in legislation (e.g. Convention of Biological Diversity), although standardised genetic diversity assessments are yet to be developed [93]. Population genomics will also enable the association of adaptive traits with genomic features, informing captive breeding and genetic rescue efforts [94], as well as the identification of limits to evolutionary rescue [95]. The dense genomic sampling will also permit studies of species delimitation, including within **cryptic species complexes** [96,97], improving our understanding of biodiversity at the continental scale [98] and inform conservation measures **(Figure 3Q)**. The use of environmental DNA metagenomics to monitor biodiversity is rapidly growing in response to a greater appreciation of the value of ecosystem-oriented conservation. This will benefit from a suite of reference genomes, allowing more accurate tracking of patterns of species' loss as their critical environmental limits are exceeded.

Project Psyche genomes will also accelerate our ability to address societal challenges. Lepidoptera have potential for biotechnology that can be enhanced with reference genomes, as exemplified by the production of more extensible silk in the moth *B. mori* [99]. Additionally, the Greater wax moth (*Galleria mellonella*) can degrade polyethylene, a prevalent form of plastic [100] and is used as a model for immune responses against human pathogens [101] **(Figure 3N)**.

As well as critical sources of food security for many human societies [102], with potential for domestication and commercial cultivation in future food economies, some Lepidoptera are pests, causing high economic and societal costs. Genomes provide a powerful resource to monitor outbreaks, investigate the genomic basis of invasiveness and harness this understanding to provide solutions for pest management **(Figure 3M)**. Traditionally, broad-spectrum insecticides have been the main form of control. Reference genomes of pest species and their close relatives enable deciphering the mechanisms of insecticide resistance, sometimes acquired via introgression [73]. Artificial intelligence can search the genomes for novel targets for highly specific insecticides. However, given the strong impact of insecticides on biodiversity [103] and human health [104], an alternative approach is to implement lower ecological impact strategies including biocontrol and agroecology, benefitting from natural enemies of pests [105]. Likewise, microbiome and virome data [106–108] and genomic insights into how co-bionts manipulate host fitness [109,110] are opening new approaches to pest control. Another potential avenue for pest control facilitated by genomes is gene drive [111], although this is not without risks [112]. Ultimately, Project Psyche aims to work synergistically with policy makers to integrate genomics and the ecological knowledge generated by the data into sustainable pest management **(Figure 3O)**.

Concluding remarks

The human reference genome underpinned a paradigm shift in our understanding of human health and disease. Twenty-five years on from its draft assembly, the vast majority of species on Earth are yet to have a reference genome. This limits our understanding of their ecology and evolution, and how this genomic and phenotypic diversity contributes to the resilience of ecosystems and their outputs. Project Psyche will provide reference genomes for 11,000 species of Lepidoptera and propel scientific discovery and novel solutions for major societal challenges, including food production and biodiversity conservation (see **Outstanding Questions**). Reference genomes from across this entire order, about which we already have remarkable natural history and historical knowledge, will transform research, shifting single-species foci towards understanding the general principles of evolutionary mechanisms across all levels of genome organisation. We are already seeing Project Psyche inspire and empower similar collaborative projects for other taxonomic groups and geographic areas. This is central to achieving the overarching aim of the Earth Biogenome Project to sequence all eukaryotic species. With the ever-improving approaches to utilise reference genomes with methods including artificial intelligence, these reference genomes will provide a rich and vital resource for the ecological and climate emergencies challenges of our era.

Outstanding questions

- How generalisable are findings from previous genomic studies based on selected single species or genera when thousands of genomes are studied?
- How and why do lineages differ in their species richness, rates of genomic rearrangements and/or genetic diversity?
- What genomic features are associated with major evolutionary transitions, such as changing between diurnal and nocturnal lifestyles, monophagous or polyphagous feeding behaviours, or aquatic and terrestrial habitats? Studying many transitions will reveal whether specific genomic features are repeatedly involved, and whether such changes are reversible, or consistently rapid.
- How can genomics be integrated into a standardised framework of conservation management, and what are the key parameters relevant to conservation that can be obtained from genomics?
- What new approaches for sustainable pest management will be possible thanks to a comprehensive catalogue of lepidopteran genomes in Europe? A key aim for pest management is reducing the ecological impact, with more geographically and taxonomically-restricted effects.
- What new discoveries in bioengineering will be facilitated by thousands of genomes?

Box 1: Inclusivity and Accessibility

The Psyche network currently has 184 members from 34 countries **(Figure 2C)**. Inclusivity is a cornerstone of our consortium, shaping how we generate and share knowledge in lepidopteran genomics. Our commitment to open data ensures transparency and access to genomic resources for diverse stakeholders, from scientists and funders to the public. To minimise the need for restrictive publishing embargoes, we will foster collaboration and fair data use agreements. We also have clear publication guidelines for using Project Psyche genomes to recognise the efforts of the Project Psyche community in studies that are made possible by these genomes (https://www.projectpsyche.org/participate/genome-analysis/). During in-person meetings, considerations are made to ensure equitable attendance and opportunities for socialising, together with accessibility of premises and caregiving responsibilities, to support a representative and inclusive community. Finally, project outputs will be openly shared and promoted through peer-reviewed articles, conferences, and media outreach in multiple languages, highlighting scientific value and public investment returns.

Gender inclusivity is another key priority, given the traditionally male-dominated landscape of entomology [113]. Although the current Psyche consortium is only 38.6% women, the project has women in key leadership roles, and we actively promote inclusive leadership with equitable decision-making and supportive work environments. These actions, aligned with the EU Gender Equality Strategy [114], aim to improve publication and funding rates for women and provide visible role models to inspire future generations. Public engagement activities further amplify diverse voices within the project including early-career scientists.

Equity also means addressing structural disparities across Europe. Partners in **European Widening Countries** often face challenges in access to infrastructure, training, and funding opportunities. These gaps are compounded by undervalued taxonomic expertise and comparatively lower genomic literacy. We aspire to address this through tailored capacity-building, including joint sampling events, mobility schemes and data hackathons. The successful establishment of a COST action “10kLepGenomes” focused around Psyche has released funding to support these collaborations and training opportunities, particularly for early career researchers. We are building a “genome native” generation of researchers who will be able to deploy genomic and genetic tools and approaches throughout their future careers. By embedding inclusivity into our structure and research - from gender equity to geographic balance and open knowledge sharing - we aim to build a collaborative model that advances biodiversity genomics and fosters a good research culture.

Box 2: Overcoming challenges

Project Psyche relies on the generosity of research teams across the continent, who voluntarily contribute to the project. However, a large network of people with diverse interests comes with the challenge of coordinating and maintaining momentum. To overcome this challenge, Project Psyche uses many communication channels, including online meetings, email, Discord, as well as in person meetings, to enable open communication and broad input. Another challenge is encouraging long-term participation and ownership. We do this through acknowledging contributions in diverse ways, including through publications and community-awards (for example, genome curator of the month). Managing expectations is also key. For example, not all collected specimens will be sequenced and timelines for genome production are not always linear. The development of the Project Psyche portal helps avoid duplication of collection efforts, and tracking of species through the genome generation pipeline.

A major challenge is the need for a dynamic checklist of European Lepidoptera that is generally accepted and implemented across sequencing databases. An updated checklist will represent a key deliverable of Project Psyche. To extend the network of collaborators and geographic areas, we are implementing further sub-hubs to temporarily store samples. However, maintaining a cold chain for samples is not always possible, especially in remote areas. We are working towards breaking the cold chain by developing new protocols that relax this requirement.

Sequencing 11,000 species will require massive upscaling of processes. As the easiest species to find and identify are sequenced, the remaining species will be challenging to collect and identify. DNA barcoding and genital determination by taxonomists will become crucial. It will also become necessary to develop protocols to extract and sequence small amounts of DNA for extremely small specimens. Methods such as Picogram input multimodal sequencing (PiMmS) [115] are promising avenues. Currently, the most time-consuming step in genome assembly remains manual curation, limiting the rate of genome generation. Alongside increasing numbers of trained curators, Project Psyche will work with other biodiversity sequencing projects to develop automated tools for curation. Another challenge is fully resolving highly repetitive genomic regions such as those on many W chromosomes. Methods to extract and sequence very long DNA molecules (>150 kb) using Oxford Nanopore technologies are being developed and are expected to help resolve such difficulties.

Acknowledgements

The writing of this manuscript was initiated during a workshop at Lund University in Sweden which was funded and organised through the 10kLepGenomes COST Action CA23122. C.J.W., M.B. and J.I.M. acknowledge funding support from the Wellcome Trust award 220540/Z/20/A ‘Wellcome Sanger Institute Quinquennial Review 2021–2026’. M.B. acknowledges funding support from Wellcome Trust grant no. 218328. J.I.M. was supported by a Branco Weiss Fellowship and a Royal Society University Research Fellowship (URF\R1\221041). K.L. was supported by the Swiss National Science Foundation grants 202869 and 10003544. M.V. was supported by Xunta de Galicia (grant ED431B 2024/23). S.K. was supported by the Bulgarian Science Fund (grant number КП-06-ПБ-21). P.N., P.M.-M. and I.K. were supported by the Czech Science Foundation (grant no. 22-35084J to P.M.-M. and 23-06455S to P.N.). I.A.D receives salary support from the CNRS. P.O.M was supported by a BBSRC Fellowship (UKRI893). We thank Franz Chandi for help in gathering the data for Figure 2C. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Declaration of interests

The authors have no interests to declare.

Glossary

B chromosomes: Non-essential chromosomes that are only found in some individuals in a population or species. The number of B chromosome copies varies from none to several.

Chromosome-level reference genome: The majority of the genome is assembled into scaffolds that represent entire chromosomes, with one scaffold per chromosome. These genomes are highly complete and have high base-level accuracy.

Chromosome territory: The region of the nucleus that is preferentially occupied by a chromosome during interphase of the cell cycle.

Cobionts: Organisms that live alongside the host such as contaminants, symbionts and parasites.

Cryptic species complex: Two or more species that have been treated as the same species, usually due to morphological similarity.

Contig: A sequence without gaps that is produced by assembling DNA reads.

DNA barcoding: An approach that enables identification of a biological specimen to a species based on the sequencing of short, standard fragments of DNA.

Scaffold: A sequence composed of two or more contigs that are joined with gaps, e.g. by evidence from Hi-C data or genetic linkage maps.

Heterogametic sex: The sex that has two different types of sex chromosomes. In most mammals, males are heterogametic, carrying an X and a Y chromosome, while females are homogametic, carrying two X chromosomes. In Lepidoptera, females tend to be the heterogametic sex, typically carrying a Z and a W, whereas males have two Z chromosomes. In some Lepidoptera species, there are multiple Z chromosomes and/or multiple or no W chromosomes.

Haplotype: Physically linked genomic variants that were inherited together on the same chromosome from a single parent.

Holocentric chromosomes: Chromosomes where centromere activity is distributed along the entire length of the chromosome as opposed to monocentric chromosomes with a single centromere.

Long read data: DNA sequence data with reads of kilobases in length, such as Pacific Biosciences (PacBio) or Oxford Nanopore Technologies (ONT) reads. These are typically used for the assembly of contigs during genome assembly.

Long-range data: DNA sequence data capturing long-range chromatin interactions within the 3D structure of the nucleus, e.g. using Hi-C sequencing. DNA regions that are close to each other in 3D space are cross-linked prior to preparing the DNA for sequencing. Short-read sequencing (e.g. Illumina sequencing) then provides two reads that are from physically close genomic regions. This data is typically used for assembling scaffolds and for studying the 3D chromatin structure and gene interactions.

Macroevolution: Patterns and processes above the species level and at a large temporal scale. Typically investigated using tools from phylogenetics, comparative studies, or modelling of traits or rates, such as diversification rates.

Microevolution: Changes in allele frequency within a population, species or closely related set of species over short periods of time. Typically, using tools from population genetics.

Topologically associating domains (TADs): Regions of the genome where DNA sequences within the domain physically interact with each other more frequently than with sequences outside the domain. TADs are thought to influence gene regulation as regulatory elements and their target genes are typically found within the same TAD. They can be identified with long-range data such as Hi-C data.

Three-dimensional (3D) chromatin architecture: Spatial organisation of the complex of DNA and its proteins (e.g. histones) within the nucleus. This 3D structure plays a crucial role in regulating gene expression and other nuclear processes. It can be assessed with Hi-C data.

European Widening Countries: European Union member states and countries associated with the Horizon Europe Programme designated as well-performing (strengthening) or under-performing (widening) in research and innovation indicators, generally characterised by stronger research institutional frameworks and support (REA 2023).

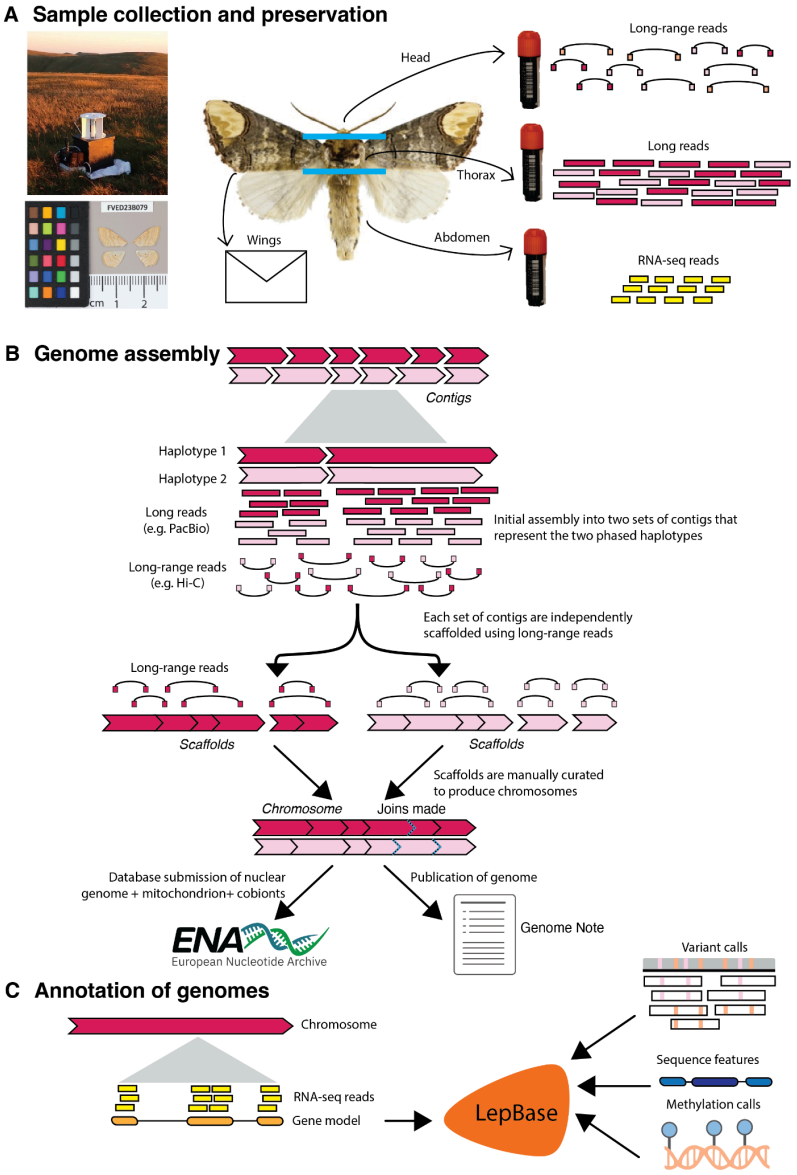

Figure 1: The Project Psyche workflow: From catching Lepidoptera to annotated reference genomes.

A. Lepidoptera are caught with traps or nets and photographed in a standardised way. For most specimens the head is used for long-range (Hi-C) sequencing, the thorax for long-read (PacBio) sequencing and the abdomen for RNA sequencing. The wings are stored in glassine envelopes.
B. PacBio reads are aligned against each other to reconstruct contiguous sequences (**contigs**) for each haplotype separately. Hi-C reads are used to identify the set of contigs belonging to the same haplotype. Because some regions are not sequenced well, e.g. repeat-rich regions, there are gaps between contigs. Hi-C reads are used to bridge those gaps and join contigs in the right order and orientation together into **scaffolds**. After manual curation, these scaffolds mostly represent entire chromosomes.
C. Protein-coding genes in the genome are annotated using species-specific RNA sequencing (RNA-seq) data and protein homology evidence. Standardised resources will also be generated for each genome, including variant calls, annotation of sequence features such as transposable elements, and 5mC methylation calls obtained from PacBio data. These annotations will all be available in LepBase [116].

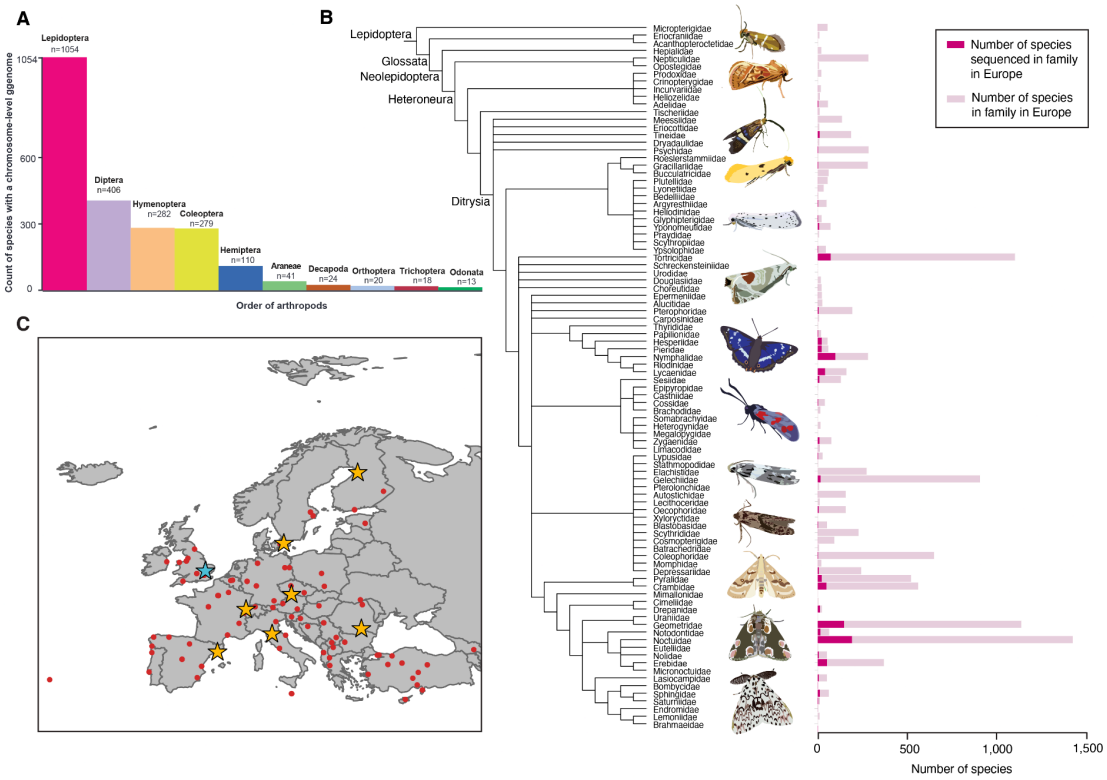

Figure 2: Sequencing the diversity of Lepidoptera found in Europe.

A. Comparison of the status of sequencing for Lepidoptera compared to the other orders within Arthropoda. Showing the ten orders with the highest numbers of chromosome-level genomes. Data derived from GoaT on 05.09.2025.
B. The tree shows family relationships between all families of Lepidoptera which have representatives in Europe. The tree structure is adapted from [54,117]. The outer ring shows a stacked bar chart of the number of species sequenced (dark pink) out of the total number per family in Europe (light pink). The figure was generated using iToL (https://itol.embl.de/). Representative species are shown alongside the phylogeny; *Micropterix aruncella* (Micropterigidae), *Triodia sylvina* (Hepialidae), *Adela croesella* (Adelidae), *Tinea trinotella* (Tineidae), *Yponomeuta malinellus* (Yponomeutidae), *Notocelia uddmanniana* (Tortricidae), *Apatura iris* (Nymphalidae), *Zygaena filipendulae* (Zygaenidae), *Recurvaria leucatella* (Gelechiidae), *Blastobasis adustella* (Blastobasidae), *Parapoynx stratiotata* (Crambidae), *Thyatira batis* (Drepanidae), *Lymantria monacha* (Erebidae).
C. Map of locations of sample collection hubs (yellow stars) and sequencing hubs (blue star). Locations of member organisations that are part of the Project Psyche consortium are represented by red circles. Note that there are also partners outside of Europe (Canada, Turkey, Georgia, Israel, USA), some of which are not depicted.

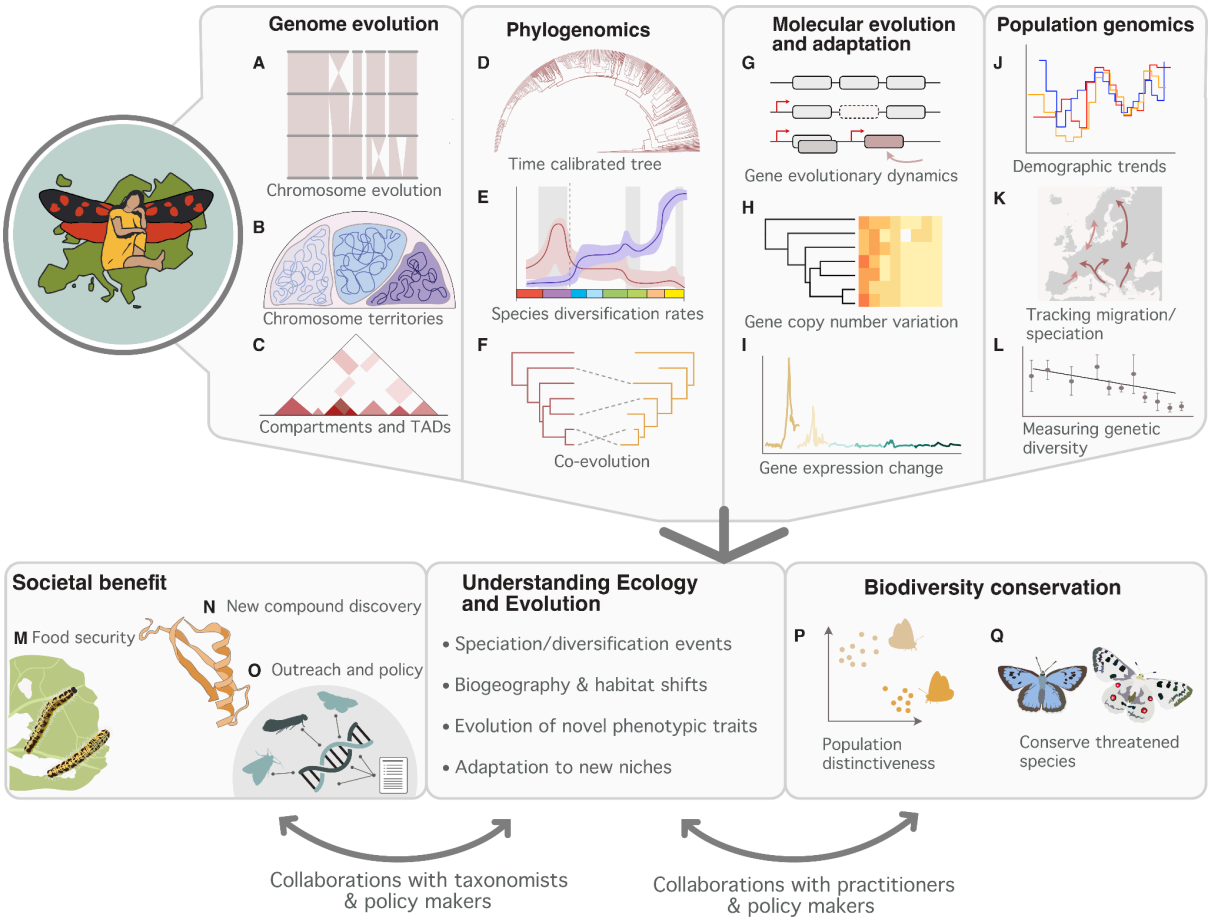

Figure 3: The potential of Psyche genomes to address evolutionary questions and impact society.

A. Evolution of chromosome structure through fusions, fissions and inversions.
B. 3D genome, chromosome territories.
C. Compartments and topologically associated domains (TADs).
D. Time-calibrated species phylogeny.
E. Diversification rates over time of two lineages.
F. Discordance between phylogenies (e.g. horizontal gene transfer, cobionts, mitochondria, hybridisation).
G. Gene duplication, loss, transfer, and gain of new regulatory elements.
H. Gene evolution/copy number variation in the context of a tree, and correlation with phenotypic traits.
I. Gene expression across different conditions/states and association with phenotypic traits.
J. Population size changes over time in three lineages.
K. Tracking species range expansion or genetic connectivity across space.
L. Correlates of genetic diversity with traits such as range size.
M. Food security. Lepidoptera are an important source of food and protein which has potential for expansion facilitated by genomics. In addition, genomics aids tracking and management of species of Lepidoptera which are major agricultural pests.
N. Discovery of new compounds for biotechnology.
O. Integrating genomics with ecology and taxonomy will feed into knowledge dissemination to the public and influence policy

P. Genetic comparisons between populations, informing conservation units and species re-introductions
Q. Conservation case studies, Large Blue and Apollo butterflies

References

1. Van Nieukerken, E.J. *et al.* (2011) Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q.(Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa* 3148, 212–221
2. Kawahara, A.Y. *et al.* (2019) Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc. Natl. Acad. Sci. U. S. A.* 116, 22657–22663
3. Misof, B. *et al.* (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767
4. Wahlberg, N. *et al.* (2013) Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). *PLoS One* 8, e80875
5. van Swaay, C.A.M. *et al.* (2022) *European Grassland Butterfly Indicator 1990-2020: Technical report*
6. Metcalf, R.L. and Luckmann, W.H., eds. (1994) *Introduction to insect pest management*, (3rd edn), John Wiley & Sons
7. Paini, D.R. *et al.* (2016) Global threat to agriculture from invasive species. *Proc. Natl. Acad. Sci. U. S. A.* 113, 7575–7579
8. *Heliconius* Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487, 94–98
9. Hill, G.M. *et al.* (2021) Climate change effects on animal ecology: butterflies and moths as a case study. *Biol. Rev. Camb. Philos. Soc.* 96, 2113–2126
10. García-Berro, A. *et al.* (2023) Migratory behaviour is positively associated with genetic diversity in butterflies. *Mol. Ecol.* 32, 560–574
11. Seibold, S. *et al.* (2019) Arthropod decline in grasslands and forests is associated with landscape-level drivers. *Nature* 574, 671–674
12. Addink, W. *et al.* (2019) DiSSCo as a new regional model for Scientific Collections in Europe. *Biodivers. Inf. Sci. Stand.* 3
13. Lewin, H.A. *et al.* (2022) The Earth BioGenome Project 2020: Starting the clock. *Proc. Natl. Acad. Sci. U. S. A.* 119
14. Melters, D.P. *et al.* (2012) Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Res.* 20, 579–593
15. Liao, W.-W. *et al.* (2022) A draft human pangenome reference. *Nature* 617, 312–324
16. Darwin Tree of Life Project Consortium, *et al.* (2022) Sequence locally, think globally: The Darwin Tree of Life Project. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2115642118
17. Mc Cartney, A.M. *et al.* (2024) The European Reference Genome Atlas: piloting a decentralised approach to equitable biodiversity genomics. *NPJ Biodivers* 3, 28
18. (2002) Lepiforum: Bestimmung von Schmetterlingen (Lepidoptera) und ihren präimaginalstadien[Online]. Available: http://www.lepiforum.de
19. Hochkirch, A. *et al.* (2022) European Red List of Insect Taxonomists. *Luxembourg: Publication Office of the European Union.*
20. Benham, P.M. and Bowie, R.C. (2022) Natural history collections as a resource for conservation genomics: understanding the past to preserve the future. *J. Hered.* DOI: 10.1093/jhered/esac066
21. Letsch, H. *et al.* (2025) Type genomics: a Framework for integrating Genomic Data into Biodiversity and Taxonomic research. *Syst. Biol.*
22. Crowley, L. *et al.* (2023) A sampling strategy for genome sequencing the British terrestrial arthropod fauna. *Wellcome Open Res.* 8, 123
23. Lawniczak, M.K.N. *et al.* (2025) Best-practice guidance for Earth BioGenome Project sample collection and processing: progress and challenges in biodiverse reference

genome creation. *Gigascience* 14, giaf041
24. Hartley, M. *et al.* (2022) The BioImage Archive - building a home for life-sciences microscopy data. *J. Mol. Biol.* 434, 167505
25. Challis, R. *et al.* (2023) Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. *Wellcome Open Research* 8
26. Vancaester, E. and Blaxter, M. (2023) Phylogenomic analysis of *Wolbachia* genomes from the Darwin Tree of Life biodiversity genomics project. *PLoS Biol.* 21, e3001972
27. Weber, C.C. (2024) Disentangling cobionts and contamination in long-read genomic data using sequence composition. *G3* 14, jkae187
28. Bortoluzzi, C. *et al.* (2023) Lepidoptera genomics based on 88 chromosomal reference sequences informs population genetic parameters for conservation *bioRxiv*, 2023.04.14.536868
29. Lawniczak, M.K.N. *et al.* (2022) Standards recommendations for the Earth BioGenome Project. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2115639118
30. Harrison, P.W. *et al.* (2021) The European Nucleotide Archive in 2020. *Nucleic Acids Res.* 49, D82–D85
31. Ratnasingham, S. and Hebert, P.D.N. (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org): BARCODING. *Mol. Ecol. Notes* 7, 355–364
32. Robinson, R. (1971) *Lepidoptera Genetics*, Pergamon Press
33. Makino, S. (1951) An atlas of the chromosome numbers in animals. *An atlas of the chromosome numbers in animals.* Available at https://www.cabdirect.org/cabdirect/abstract/19520100513
34. Wright, C.J. *et al.* (2024) Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nat Ecol Evol* 8, 777–790
35. Wright, C.J. *et al.* (2025) The 229 chromosomes of the Atlas blue butterfly reveal rules constraining chromosome evolution in Lepidoptera. *bioRxiv*, 2025.03. 14.642909
36. Senaratne, A.P. *et al.* (2021) Formation of the CenH3-Deficient Holocentromere in Lepidoptera Avoids Active Chromatin. *Curr. Biol.* 31, 173–181.e7
37. Cortes-Silva, N. *et al.* (2020) CenH3-independent kinetochore assembly in Lepidoptera requires CCAN, including CENP-T. *Curr. Biol.* 30, 561–572.e10
38. Rosin, L.F. *et al.* (2021) Oligopaint DNA FISH reveals telomere-based meiotic pairing dynamics in the silkworm, *Bombyx mori*. *PLoS Genet.* 17, e1009700
39. Hockens, C. *et al.* (2024) Chromosome segregation during spermatogenesis occurs through a unique center-kinetic mechanism in holocentric moth species. *PLoS Genet.* 20, e1011329
40. Fraïsse, C. *et al.* (2017) The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W. *Nat. Commun.* 8, 1486
41. Han, M.-J. *et al.* (2024) Multiple independent origins of the female W chromosome in moths and butterflies. *Sci Adv* 10, eadm9851
42. Dalíková, M. *et al.* (2017) New insights into the evolution of the W chromosome in Lepidoptera. *J. Hered.* 108, 709–719
43. Decroly, T. *et al.* (2024) Rewinding the ratchet: rare recombination locally rescues neo-W degeneration and generates plateaus of sex-chromosome divergence. *Mol. Biol. Evol.* DOI: 10.1093/molbev/msae124
44. Yoshido, A. *et al.* (2020) Evolution of multiple sex-chromosomes associated with dynamic genome reshuffling in *Leptidea* wood-white butterflies. *Heredity* 125, 138–154
45. Turner, J.R. and Sheppard, P.M. (1975) Absence of crossing-over in female butterflies (*Heliconius*). *Heredity* 34 Part 2, 265–269
46. Van't Hof, A.E. *et al.* (2025) Sex determination in moths and butterflies: Masculinizer as key player. *Curr. Opin. Insect Sci.* 70, 101375
47. Kiuchi, T. *et al.* (2014) A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature* 509, 633–636

48. Yoshido, A. and Marec, F. (2023) Deviations in the Z:A ratio disrupt sexual development in the eri silkmoth, *Samia cynthia ricini. Genetics* 224, iyad023
49. Van't Hof, A.E. *et al.* (2024) Zygosity-based sex determination in a butterfly drives hypervariability of *Masculinizer*. *Sci Adv* 10, eadj6979
50. Rosin, L.F. *et al.* (2022) Dosage compensation in *Bombyx mori* is achieved by partial repression of both Z chromosomes in males. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2113374119
51. Gil, J. *et al.* (2023) Unique territorial and sub-chromosomal organization revealed in the holocentric moth *Bombyx mori*. *bioRxiv* DOI: 10.1101/2023.09.14.557757
52. Rowley, M.J. and Corces, V.G. (2018) Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800
53. Condamine, F. *et al.* (2016) Global patterns of insect diversification: towards a reconciliation of fossil and molecular evidence? *Sci. Rep.* 6
54. Mitter, C. *et al.* (2017) Phylogeny and evolution of Lepidoptera. *Annu. Rev. Entomol.* 62, 265–283
55. De-Kayne, R. *et al.* (2025) Why do some lineages radiate while others do not? Perspectives for future research on adaptive radiations. *Cold Spring Harb. Perspect. Biol.* 17, a041448
56. Vellend, M. (2005) Species diversity and genetic diversity: parallel processes and correlated patterns. *Am. Nat.* 166, 199–215
57. Choi, J.Y. *et al.* (2021) Ancestral polymorphisms shape the adaptive radiation of Metrosideros across the Hawaiian Islands. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2023801118
58. Augustijnen, H. *et al.* (2024) A macroevolutionary role for chromosomal fusion and fission in *Erebia* butterflies. *Sci. Adv.* 10, eadl0989
59. Peñalba, J.V. *et al.* (2024) The role of hybridization in species formation and persistence. *Cold Spring Harb. Perspect. Biol.* 16, a041445
60. Serrato-Capuchina, A. and Matute, D.R. (2018) The role of transposable elements in speciation. *Genes (Basel)* 9, 254
61. Bista, I. *et al.* (2023) Genomics of cold adaptations in the Antarctic notothenioid fish radiation. *Nat. Commun.* 14, 3412
62. Marques, D.A. *et al.* (2019) A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* 34, 531–544
63. White, M.J.D. (1978) Modes of speciation.
64. Rieseberg, L.H. (2001) Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16, 351–358
65. McClure, M. *et al.* (2017) Heterozygosity and chain multivalents during meiosis illustrate ongoing evolution as a result of multiple holokinetic chromosome fusions in the genus Melinaea (Lepidoptera, Nymphalidae). *Cytogenetic and Genome Research* 153, 213–222
66. Lucek, K. *et al.* (2022) A holocentric twist to chromosomal speciation? *Trends Ecol. Evol.* 37, 655–662
67. Edger, P.P. *et al.* (2015) The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U. S. A.* 112, 8362–8366
68. Wheat, C.W. *et al.* (2007) The genetic basis of a plant-insect coevolutionary key innovation. *Proc. Natl. Acad. Sci. U. S. A.* 104, 20427–20431
69. Brien, M.N. *et al.* (2023) Colour polymorphism associated with a gene duplication in male wood tiger moths. *Elife* 12
70. Van't Hof, A.E. *et al.* (2016) The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534, 102–105
71. Tunström, K. *et al.* (2023) Evidence for a single, ancient origin of a genus-wide alternative life history strategy. *Sci. Adv.* 9, eabq3713
72. Walker, A.A. *et al.* (2023) Horizontal gene transfer underlies the painful stings of asp

caterpillars (Lepidoptera: Megalopygidae). *Proc. Natl. Acad. Sci. U. S. A.* 120, e2305871120
73. Valencia-Montoya, W.A. *et al.* (2020) Adaptive Introgression across Semipermeable Species Boundaries between Local *Helicoverpa zea* and Invasive *Helicoverpa armigera* Moths. *Mol. Biol. Evol.* 37, 2568–2583
74. Weisman, C.M. *et al.* (2022) Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes. *Curr. Biol.* 32, 2632–2639.e2
75. Blount, Z.D. *et al.* (2018) Contingency and determinism in evolution: Replaying life's tape. *Science* 362, eaam5979
76. Gould, S.J. (1989) Wonderful life: the Burgess Shale and the nature of history. at https://bookreadfree.com/23089/619674
77. Bohutínská, M. and Peichel, C.L. (2024) Divergence time shapes gene reuse during repeated adaptation. *Trends Ecol. Evol.* 39, 396–407
78. Nadeau, N.J. (2016) Genes controlling mimetic colour pattern variation in butterflies. *Curr. Opin. Insect Sci.* 17, 24–31
79. Mulhair, P.O. *et al.* (2023) Diversity, duplication, and genomic organization of homeobox genes in Lepidoptera. *Genome Res.* 33, 32–44
80. Pringle, E.G. *et al.* (2007) Synteny and chromosome evolution in the lepidoptera: evidence from mapping in *Heliconius melpomene*. *Genetics* 177, 417–426
81. Fandino, R.A. *et al.* (2024) The ivory lncRNA regulates seasonal color patterns in buckeye butterflies. *Proc. Natl. Acad. Sci. U. S. A.* 121, e2403426121
82. Livraghi, L. *et al.* (2024) A long noncoding RNA at the cortex locus controls adaptive coloration in butterflies. *Proc. Natl. Acad. Sci. U. S. A.* 121, e2403326121
83. Tian, S. *et al.* (2024) A microRNA is the effector gene of a classic evolutionary hotspot locus. *Science* 386, 1135–1141
84. Marlétaz, F. *et al.* (2023) The little skate genome and the evolutionary emergence of wing-like fins. *Nature* 616, 495–503
85. M Real, F. *et al.* (2020) The mole genome reveals regulatory rearrangements associated with adaptive intersexuality. *Science* 370, 208–214
86. Lupiáñez, D.G. *et al.* (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012–1025
87. Duplouy, A. and Hornett, E.A. (2018) Uncovering the hidden players in Lepidoptera biology: the heritable microbial endosymbionts. *PeerJ* 6, e4629
88. Jiggins, F.M. *et al.* (2000) The butterfly *Danaus chrysippus* is infected by a male-killing Spiroplasma bacterium. *Parasitology* 120 (Pt 5), 439–446
89. Gavotte, L. *et al.* (2007) A Survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. *Mol. Biol. Evol.* 24, 427–435
90. Bridle, J. and Hoffmann, A. (2022) Understanding the biology of species' ranges: when and how does evolution change the rules of ecological engagement? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 377, 20210027
91. de Jong, M. *et al.* (2023) Rapid evolution of novel biotic interactions in the UK Brown Argus butterfly uses genomic variation from across its geographical range. *Mol. Ecol.* 32, 5742–5756
92. Liddell, E. *et al.* (2020) Evaluating the use of risk assessment frameworks in the identification of population units for biodiversity conservation. *Wildl. Res.* 47, 208
93. Buzan, E. *et al.* (2025) Biodiversity genomics research practices require harmonising to meet stakeholder needs in conservation. *Mol. Ecol.*
94. Halford, G. *et al.* (2025) Genomic monitoring of a reintroduced butterfly uncovers contrasting founder lineage survival. *Evol. Appl.* 18, e70074
95. Chevin, L.M. and Bridle, J. (2025) Impacts of limits to adaptation on population and community persistence in a changing environment. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 380, 20230322

96. Lucek, K. *et al.* (2024) Towards a genomic resolution of the *Phengaris alcon* species complex. *Conserv. Genet.* 25, 823–834
97. Jospin, A. *et al.* (2023) Genomic evidence for three distinct species in the *Erebia manto* complex in Central Europe (Lepidoptera, Nymphalidae). *Conserv. Genet.* 24, 293–304
98. Vodă, R. *et al.* (2015) Cryptic matters: overlooked species generate most butterfly beta-diversity. *Ecography (Cop.)* 38, 405–409
99. Xu, J. *et al.* (2018) Mass spider silk production through targeted gene replacement in *Bombyx mori*. *Proc. Natl. Acad. Sci. U. S. A.* 115, 8757–8762
100. Bombelli, P. *et al.* (2017) Polyethylene bio-degradation by caterpillars of the wax moth Galleria mellonella. *Curr. Biol.* 27, R292–R293
101. Pereira, T.C. *et al.* (2018) Recent advances in the use of *Galleria mellonella* model to study immune responses against human pathogens. *J. Fungi (Basel)* 4, 128
102. Siddiqui, S.A. *et al.* (2023) Edible Lepidoptera as human foods – a comprehensive review. *J. Insects Food Feed* 10, 25–49
103. Rigal, S. *et al.* (2023) Farmland practices are driving bird population decline across Europe. *Proc. Natl. Acad. Sci. U. S. A.* 120, e2216573120
104. Cavalier, H. *et al.* (2023) Exposures to pesticides and risk of cancer: Evaluation of recent epidemiological evidence in humans and paths forward. *Int. J. Cancer* 152, 879–912
105. Goulet, F. *et al.* (2023) Moving beyond pesticides: Exploring alternatives for a changing food system. *Environ. Sci. Policy* 147, 177–187
106. Niu, D. *et al.* (2024) Physiological and molecular mechanisms of Lepidopteran insects: Genomic insights and applications of genome editing for future research. *Int. J. Mol. Sci.* 25, 12360
107. Tay, W.T. *et al.* (2023) *Spodoptera frugiperda*: Ecology, evolution, and management options of an invasive species. *Annu. Rev. Entomol.* 68, 299–317
108. Amado, D. *et al.* (2025) The genetic architecture of resistance to flubendiamide insecticide in *Helicoverpa armigera* (Hübner). *PLoS One* 20, e0318154
109. Hoffmann, A.A. *et al.* (2015) Wolbachia strains for disease control: ecological and evolutionary considerations. *Evol. Appl.* 8, 751–768
110. Martinez, J. *et al.* (2014) Symbionts commonly provide broad spectrum resistance to viruses in insects: a comparative analysis of Wolbachia strains. *PLoS Pathog.* 10, e1004369
111. Holman, L. (2019) Evolutionary simulations of Z-linked suppression gene drives. *Proc. Biol. Sci.* 286, 20191070
112. Courtier-Orgogozo, V. *et al.* (2020) Evaluating the probability of CRISPR-based gene drive contaminating another species. *Evol. Appl.* 13, 1888–1905
113. Walker, K.A. (2018) Gender gap in professional entomology: Women are underrepresented in academia and the U.S. government. *Ann. Entomol. Soc. Am.* 111, 355–362
114. corporate-body. RTD:Directorate-General for Research and Innovation (2025) *She figures 2024*, Publications Office of the European Union
115. Laumer, C. (2023) Picogram input multimodal sequencing (PiMmS). at https://www.protocols.io/view/picogram-input-multimodal-sequencing-pimms-cfaitic
116. Challis, R.J. *et al.* (2016) Lepbase: the Lepidopteran genome database. *bioRxiv*, 056994
117. Rota, J. *et al.* (2022) The unresolved phylogenomic tree of butterflies and moths (Lepidoptera): Assessing the potential causes and consequences. *Syst. Entomol.* 47, 531–550