CellPress
OPEN ACCESS

Trends in
Ecology & Evolution

Feature Review

Project Psyche: 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[21,22] Mark Blaxter [1], and Joana I. Meier [1,20,*]

Project Psyche is a transnational initiative to generate and study chromosome-level reference genomes of all ~11 000 species of Lepidoptera (butterflies and moths) found in Europe. Here, we describe the decentralised network of collection and sequencing hubs that has enabled rapid progress, the standardised protocols for sampling and sequencing, and the collaborative framework for data analysis. With over 1000 species already sequenced, Lepidoptera are at the forefront of biodiversity genomics with the most reference genomes of any eukaryotic order. The completed pan-European catalogue of openly accessible lepidopteran genomes will transform our understanding of evolution and ecology, inform conservation, and foster advances in management of pests and invasive species. We highlight research areas that will benefit from this large-scale genome dataset.

Lepidoptera as a window into ecology and evolution

With ~160 000 described species [1], Lepidoptera encompass ~10% of all known eukaryotic species. This megadiversity is the product of more than 230 million years of evolution [2–4], with numerous species central to many pivotal ecological functions, including herbivory and pollination. Lepidoptera include flagship species for conservation and the protection of entire ecosystems [5]. Some species are 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 and entomology. This long-term interest in butterflies and moths has made them one of the best-studied taxonomic orders. Consequently, we possess a 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** (see Glossary) studies using short-read resequencing, RNA, or epigenetics data that are mapped to the reference genome, while providing connections to **macroevolutionary** studies via

Highlights

Lepidoptera are key to ecosystem functioning as herbivores, pollinators, and prey, and are important indicators of ecosystem health. They also include many species of economic and agricultural importance. Given the increasingly alarming decline of insects, understanding key challenges such as population fragmentation and factors explaining ecological resilience are urgently needed.

Building on other large-scale biodiversity sequencing initiatives under the Earth Biogenome Project umbrella, Project Psyche aims to generate unprecedented genomic resources for Lepidoptera.

The Project Psyche community encompasses diverse researchers, amateur lepidopterists, conservation practitioners, and industry experts united by a common vision of the importance of lepidopteran genomes.

With 1000 high-quality genomes already generated in a standardised manner, Project Psyche is fuelling diverse research areas, including comparative genomics, phylogenomics, conservation, molecular evolution, and population genomics.

Sequencing all 11 000 species will set a sound foundation for genomics and greatly foster monitoring of all Lepidoptera in Europe, empowering effective biodiversity management and policy, locally and globally.

[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

https://doi.org/10.1016/j.tree.2025.10.007

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 decreasing sequencing costs, generating thousands of chromosome-level genomes is now feasible. Indeed, the Earth Biogenome Project (EBP) aims to generate reference genomes for 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 ~11 000 species of Lepidoptera occurring in Europe. The project was named after the Greek goddess of the soul, Psyche, depicted with butterfly wings, and after the ancient Greek word for butterfly itself. 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 typically small (~500 Mb) and their **holocentric chromosomes** [14] lack large, often difficult-to-assemble localised centromeric regions [15]. These features facilitate 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], which 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 a deeper understanding of biodiversity through the exploration of general principles in evolution 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 **cobiont** 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.

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 (Boxes 1–3). Currently, 11 665 species of Lepidoptera have been recorded in Europe (including migrants, invasive, and vagrant species) (http://www.lepiforum.de). Their taxonomy 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 upon taxonomic expertise to sample and identify specimens. In turn, reference genomes, combined with other types of data, including morphological, chemical, and biogeographical information, help to resolve taxonomic uncertainties.

Due to the declining number of taxonomists [18], 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, ecological, 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 [19,20] provides further historical and contemporary data to complement field observations and genomics data.

[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, Firenze, Italy
[8]"Grigore Antipa" National Museum of Natural History, Bucharest, Romania
[9]Department of Zoology, Stockholm University, 10691 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, UMR3664 Nuclear Dynamics Unit, 75005 Paris, France
[19]Department of Ecology and Vertebrate Zoology, Faculty of Biology and Environmental Protection, University of Lodz, Łódź, Poland
[20]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, United Kingdom
[21]The Project Psyche Community: https://zenodo.org/records/17314788
[22]Additional members of the Project Psyche Community: Giada Zucco: CREA, Research Centre for Forestry and Wood; Department of Agriculture,

Box 1. Inclusivity and accessibility

The Project Psyche network currently has 184 members from 34 countries (see Figure 1C in the main text). 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 foster collaboration and fair data use agreements. We also have clear authorship guidelines to recognise the efforts of the Project Psyche community in studies that are made possible by these genomes [136]. 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 [137]. 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 [138], 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 Project Psyche provides funding to support these collaborations and training opportunities, particularly for early career researchers and researchers from European Widening Countries. We are building a genome-fluent 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, strengthens engagement with diverse stakeholders, and fosters a good research culture.

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.

Box 2. Establishing Project Psyche

Project Psyche was an idea born out of the success of pioneering lepidopteran genome projects, and the growth of genomics-enabled research on Lepidoptera. The communities that grew up around silk moth (*Bombyx mori*) developmental biology, the genetics of mimicry and speciation in *Heliconius* butterflies, and the migratory behaviour of monarch butterflies (*Danaus plexippus*) showed that Lepidoptera were amenable and tractable genomics models. The DToL project, established in 2019 to sequence the genomes of the eukaryotic biota of Ireland and Britain [16], had an early focus on Lepidoptera and this led to a rich network of collaborations across Europe.

The basic structure of Phase 1 of Project Psyche built on the rich experience from DToL. This identified the benefit of having a few, engaged sample collection hubs who build strong relationships with an expert taxonomic community. These hubs would acquire skills in the technical processes of sample preservation and data recording. Project Psyche hubs currently organise their activities alongside their other commitments without dedicated funding. Sequencing a single lepidopteran reference genome currently costs about €2500 (€1500 for PacBio, €500 for Hi-C, €300 for RNA sequencing, and €200 compute costs), excluding salaries. We chose to sequence through the Wellcome Sanger Institute during Phase 1, as their funding model made it possible to pivot funding already won to generate genomes for up to 2000 species and in-house sequencing reduced the prices. The Wellcome Sanger Institute is a research-focused genomics institute that includes research on Lepidoptera. It also plays a key role in driving several large-scale genomic projects, including DToL and Project Psyche. It thus already had an established infrastructure including teams dedicated to project management, sample management, legal and compliance, laboratory work, sequencing, genome assembly, genome curation, informatics, and digital platform development.

Once formulated, the idea of Project Psyche was communicated at international conferences on Lepidoptera. It received enthusiastic support and the seven sample collection hubs quickly identified themselves. A series of monthly online meetings was established, where all aspects of the project were collectively discussed and decided from how to collect samples to the publication code of conduct. The monthly meetings now focus more on working towards publications and acquiring future funding to support the establishment of additional sequencing and sampling hubs. The project continues to encourage and welcome new members of the community and expands activities through the 10kLepGenomes COST Action (https://10klepgenomes.eu, https://www.cost.eu/actions/CA23122/).

Mediterranean University of Reggio Calabria, Rende, Italy/Reggio Calabria, Italy.
Milica Zlatkovic: Institute of Lowland Forestry and Environment (ILFE), University of Novi Sad, Antona Cehova 13d, 21000 Novi Sad, Serbia.
Marcin Wiorek: Institute of Systematics and Evolution of Animals Polish Academy of Sciences, Sławkowska 17, 31-016 Kraków, Poland.
Martin Wiemers: Senckenberg German Entomological Institute, Müncheberg, Germany.
Claudia C. Weber: Tree of Life Programme, Wellcome Sanger Institute Cambridge, United Kingdom. Pavel Vrba: Biology Centre CAS, Institute of Entomology, Ceske Budejovice, Czechia.
Vladimir Visacki: Institute of Lowland Forestry and Environment (ILFE), University of Novi Sad Antona Cehova 13d, 21000 Novi Sad, Serbia.
Jose Vicente Pérez Santa Rita: Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Valencia, Spain.
Mikko Vallinmäki: Zoological Museum of Oulu University, Oulu, Finland.
Başak Ulaşlı: Hatay Mustafa Kemal Unversity,Faculty of Agriculture, Department of Plant Protection, Hatay, Türkiye.
Eric Toro-Delgado: Institut de Biologia Evolutiva, CSIC - Universitat Pompeu Fabra, Barcelona, Spain.
Gerhard G. Thallinger: Graz University of Technology, Institute of Biomedical Informatics, Graz, Austria.
Alena Suchackova Bartonova: Senckenberg German Entomological Institute & Biology Centre CAS, Müncheberg, Germany.
Jenny Stewart: Queen Mary University of London, London, United Kingdom. Clara Spilker: Senckenberg Deutsches Entomologisches Institut Müncheberg, Eberswalder Straße 90, 15374 Müncheberg, Germany.
Loukia Spilani: Institut de Biologia Evolutiva, CSIC - Universitat Pompeu Fabra, Barcelona, Spain.
Dominik Schmitz: Senckenberg German Entomological Institute, Müncheberg, Germany.
Benjamin Schattanek-Wiesmar: Tiroler Landesmuseen, Sammlungs- und Forschungszentrum,Krajnc-Straße 1, 6060, Austria.
Petra Schattanek-Wiesmair: Tiroler Landesmuseen, Sammlungs- und Forschungszentrum,Krajnc-Straße 1, 6060 Hall in Tirol, Austria.
Stefano Scalercio: CREA, Research Centre for Forestry and Wood, Rende, Italy.

Box 3. Overcoming challenges

Project Psyche relies on the generosity of research teams across Europe, who voluntarily contribute to the project. Such a large network of people with diverse interests comes with the challenge of coordinating and maintaining momentum. To overcome this challenge, Project Psyche uses diverse 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 (e.g., genome curator of the month). Managing expectations is also key. For example, not all collected specimens will be sequenced and timelines for genome production vary. The development of the Project Psyche portal (https://psyche.tol.sanger.ac.uk/) reduces duplication of collection efforts.

Sequencing 11 000 species will require massive upscaling of processes and the acquisition of additional funding. A major challenge is the need for a dynamic checklist of European Lepidoptera that is generally accepted and implemented across sequencing databases – a key deliverable of Project Psyche. As the easiest species are sequenced, the remaining species will be challenging to collect and identify. To expand the network of collectors, 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 while maintaining high-quality RNA and DNA with intact nuclei and 3D chromatin structure. To identify challenging species, DNA barcoding and genital determination by taxonomists will become crucial. For extremely small specimens, it will also become necessary to develop protocols such as Picogram input multimodal sequencing (PiMmS) [139]. Currently, the most time-consuming step in genome assembly remains manual genome curation. 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 (e.g., on W chromosomes). Method developments to sequence very long DNA molecules (>150 kb) using Oxford Nanopore technologies (ONT) might help curation efforts for specific species with particularly repeat-rich W chromosomes.

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 highly localised species or to difficult species complexes, requiring specialised taxonomic knowledge and in some cases, additionally **DNA barcoding** for identification. Many of the easily identifiable and widespread taxa are included in the first 1000 genomes available as of September 2025 (Figure 1A). Future collecting trips will target regions with high endemicity, such as the Iberian, Italian and Balkan Peninsulas, the Alps, and the Mediterranean islands. Finally, challenging species complexes will be resolved by working closely with taxonomists, and complemented by population genomics to refine delimitation of species, and inform an updated European checklist.

Sequencing Lepidoptera at scale

Project Psyche is committed to ethical and sustainable sampling practices [17]. All sample collection strictly follows local and national regulations, also aiming to minimise the impact on protected areas and species. Sample collectors need to be aware of specific requirements in relation to local communities and indigenous people. Importantly, the Sámi people manage land in the Nordic Countries [21] and many areas are managed by local communities across Europe. Following recommendations by EBP, Project Psyche is guided by the Kunming–Montreal Global Biodiversity Framework to recognise and include indigenous peoples and local communities in efforts to protect and restore biodiversity [22–24]. Project Psyche is committed towards benefit-sharing across Europe and strives for equitable, inclusive involvement of communities across the activities of the project. The transfer of samples to the regional hubs and subsequent sequencing locations further adheres to international regulations including Nagoya and Convention on International Trade in Endangered Species (CITES).

We use a standardised methodology from sample collection to genome generation described in detail in [25] (Figure 2), building on protocols developed by the DToL project [26]. In brief, sampling aims for one to five females per species 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. Freshly killed specimens are dissected on dry ice and snap-

Gábor Sramkó Sramkó: Evolutionary Genomics Research Group, Department of Botany, University of Debrecen, Egyetem tér 1., 4032 Debrecen,Hungary.
Dejan V. Stojanovic: Institute of Lowland Forestry and Environment (ILFE), University of Novi Sad, Antona Cehova 13d, 21000 Novi Sad, Serbia,.
Cibele G. Sotero-Caio: Tree of Life Programme, Wellcome Sanger Institute, Cambridge, United Kingdom.
Jadranka Rota: Biological Museum, Department of Biology, Lund University, Lund, Sweden.
Mar Repullés: Biology Centre CAS, Institute of Entomology; Faculty of science, University of South Bohemia (USB), Ceske Budejovice, Branisovska 31, Czechia.
Łukasz Przybyłowicz: Institute of Systematics and Evolution of Animals Polish Academy of Sciences, Slawkowska 17; 31-016 Kraków, Poland.
Pavel Potocký: Biology Centre CAS, Institute of Entomology, Ceske Budejovice, Branisovska 31, Czechia.
Jaakko Pohjoismäki: Department of Environmental and Biological Sciences, University of Eastern Finland Yliopistokatu 7, 80101 Joensuu, Finland.
Erki Õunap: Institute of Ecology and Earth Sciences, University of Tartu, J. Liivi 2, 50409 Tartu, Estonia; Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Fr.R.Kreutzwaldi 1, 51006 Tartu, Estonia.
Antonio S. Ortiz: Universidad de Murcia, Murcia, Spain.
Rayner Nunez: LIB- Koenig Museum, Bonn, Germany.
Vazrick Nazari: University of Padova, Padova, Italy.
Ana Nahirnić-Beshkova: National Museum of Natural History, Bulgarian Academy of Sciences, Sofia, Bulgaria.
Tatiana Moreira: MNHNC - National Museum of Natural History and Science | CE3C, University of Lisbon | Christopher Wheat Lab, Stockholm University, Lisbon, Portugal | Stockholm, Portugal.
Marco Meyer: Tree of Life Programme, Wellcome Sanger Institute Cambridge, UK.
Mattia Menchetti: Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Spain.
Duane McKenna: University of Memphis Department of Biological Sciences and Center for Biodiversity and Global Change Research, Memphis, TN, 38152, United States.
Valéria Marques: Institut de Biologia Evolutiva, CSIC - Universitat Pompeu Fabra, Barcelona, Spain.
Konrad Lohse: Institute of Ecology and

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 3). Typically, the head is used for **long-range sequencing** (currently Hi-C with Illumina short-read sequencing), the thorax for **long-read sequencing** (currently with Pacific Biosciences, PacBio), and the abdomen or another individual for gene annotation using RNA sequencing [27] (Figure 2A). For some specimens, legs or genitalia may also be removed to confirm identification with DNA barcoding or diagnostic morphological characters, respectively. 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) [28]. To facilitate coordinated sample collections both within Project Psyche and with other biodiversity genomics projects, we developed a portal (https://psyche.tol.sanger.ac.uk/) that shows the collection and sequencing status for each species. Our species list is also integrated into the Genomes on a Tree (GoaT) database (https://goat.genomehubs.org) [29], which is used to coordinate ongoing reference genome sequencing efforts across many projects, to minimise accidental duplication of genome sequencing. GoaT also estimates 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 [30,31]. Their genomes are assembled separately from the lepidopteran mitochondrial and nuclear genome. Next, the nuclear genome is assembled and **scaffolded** from the **long-read** (PacBio) and **long-range** (Hi-C) data (Figure 2B), using EBP best-practice approaches. The high heterozygosity of lepidopteran genomes [32] allows the use of 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 ([33]; www.earthbiogenome.org). This time-consuming manual task [34] 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. Once complete, all components (raw data, both nuclear haplotypes, mitochondrial and cobiont genomes), are submitted to the European Nucleotide Archive (ENA) [35]. In addition, DNA barcode sequences used to confirm specimen identification are uploaded to the Barcode of Life Data Systems (BOLD) [36]. 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 2C), 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 they meet researcher needs and facilitate collaborative science.

Strengthening the genomics community across Europe and beyond

Project Psyche has made rapid progress since launching in August 2023, having achieved its first milestone of 1000 assembled genomes and 2000 species collected (Figure S1 in the supplemental information online). The 1000 genomes include 652 species sequenced at the WSI as part of the DToL project. Similarly, species sequenced by Project Psyche also contribute to the aims of

Evolution, University of Edinburgh, Edinburgh, United Kingdom.
Lars Littmann: Centre for Palaeogenetics, Stockholm, University Svante Arrhenius väg 20C, SE-106 91 Stockholm, Sweden.
Daniel Linke: Biology Centre CAS, Institute of Entomology, Ceske Budejovice, Branisovska 31, Czechia.
Josep Lancho: Institut de Biologia Evolutiva, CSIC - Universitat Pompeu Fabra, Barcelona, Spain.
Sara La Cava: CREA, Research Centre for Forestry and Wood; Department of Biology, Ecology and Earth Sciences, University of Calabria, Rende, Italy.
Eva Kriegova: Biology Centre CAS, Institute of Entomology, Ceske Budejovice, Branisovska 32, Czechia.
Emre Keskin: Evolutionary Genetics Laboratory (eGL), Ankara University, Ankara, Türkiye.
Stefaniya Kamenova: National Museum of Natural History, Bulgarian Academy of Sciences, Sofia, Bulgaria.
Mathieu Joron: CNRS, Centre d'Ecologie Fonctionnelle et Evolutive, Montpellier, France.
Anna K. Hundsdoerfer: Senckenberg Natural History Collections, Museum of Zoology, Koenigsbruecker Landstrasse 159, D-01109 Dresden, Germany.
Noémie Hévin: Institut des Sciences de l'Évolution de Montpellier, CNRS, Montpellier, France.
Jérémy Gauthier: Naturéum - State Museum of Natural Sciences, Palais de Rumine, Place de la Riponne 6, 1005 Lausanne, Switzerland.
Madeleine Fabusova: School of Natural and Environmental Sciences, Newcastle University, Newcastle Upon Tyne, United Kingdom.
Zdenek F. Fric: Biology Centre CAS, Institute of Entomology, Ceske Budejovice, Czechia.
Paula Escuer: Biodiversity Genomics Laboratory, University of Neuchâtel, Neuchâtel, Switzerland.
Brent Emerson: Institute of Natural Products and Agrobiology, Spanish National research Council, C/Astrofísico Francisco Sánchez 3, La Laguna, Tenerife, Canary Islands, 38206, Spain.
Marianne Elias: Institut de Systématique, Evolution, Biodiversité, MNHN-CNRS-SU-EPHE-UA, 45 rue Buffon, 75005 Paris, France.
Lora Downes: Tree of Life Programme, Wellcome Sanger Institute, Cambridge, United Kingdom.
Mónica Doblas-Bajo: Department of Organismal and Systems Biology, University of Oviedo Oviedo, Spain.
Laurence Despres: Université Grenoble Alpes, Laboratoire d'Ecologie Alpine, Grenoble, France.

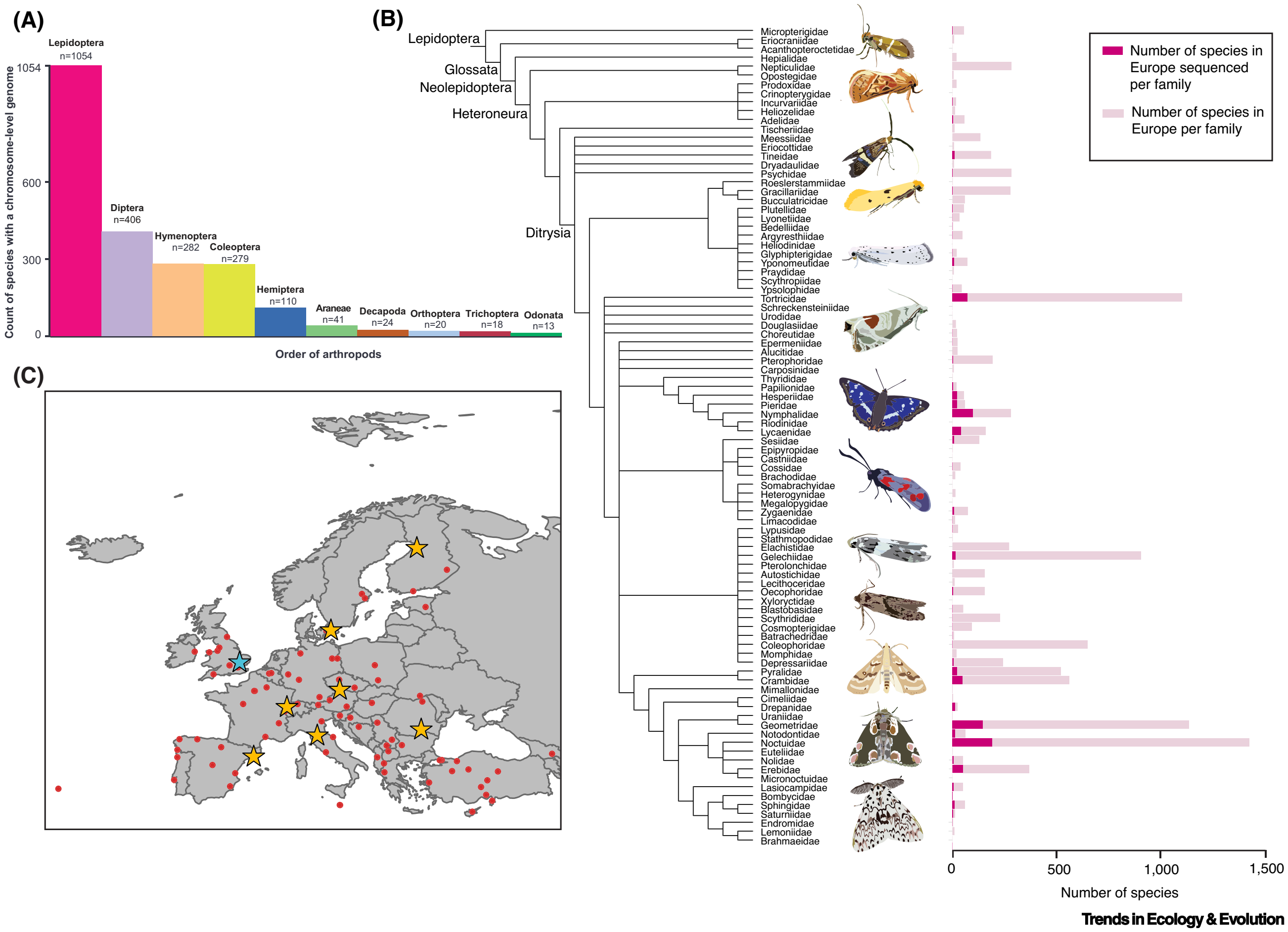


Figure 1. Sequencing the diversity of Lepidoptera found in Europe. (A) Comparison of the status of sequencing for Lepidoptera compared with the other orders within Arthropoda. Showing the ten orders with the highest numbers of chromosome-level genomes. Data derived from Genomes on a Tree (GoaT) [29] release 2025.09.26 on 26 September 2025. (B) The tree shows family relationships between all families of Lepidoptera which have representatives in Europe. The tree structure is adapted from [71,141]. 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). Data derived from GoaT [29] release 2025.09.26 on 26 September 2025. The figure was generated using iToL (https://itol.embl.de/). Representative species are shown alongside the phylogeny; white-line pollen moth (*Micropterix aruncella*, Micropterigidae), orange swift (*Triodia sylvina*, Hepialidae), lesser banded longhorn (*Adela croesella*, Adelidae), small greyish-buff moth (*Tinea trinotella*, Tineidae), apple ermine moth (*Yponomeuta malinella*, Yponomeutidae), bramble shoot moth (*Notocelia uddmanniana,* Tortricidae), purple emperor (*Apatura iris*, Nymphalidae), six-spot burnet (*Zygaena filipendulae*, Zygaenidae), lesser budmoth (*Recurvaria leucatella*, Gelechiidae), common masoner (*Blastobasis adustella*, Blastobasidae), ringed china-mark (*Parapoynx stratiotata*, Crambidae), peach blossom (*Thyatira batis*, Drepanidae), and black arches (*Lymantria monacha*, Erebidae). (C) Map of locations of sample collection hubs (yellow stars) and sequencing hub (blue star). Locations of member organisations that are part of the Project Psyche consortium are represented by red circles. Note that the Canary Islands are not illustrated, although they are included in the area studied and members of Project Psyche are located there. There are also partners outside of Europe (Canada, Turkey, Georgia, Israel, USA), some of which are not depicted.

other projects including ERGA and EBP. Many of the species sequenced by Project Psyche have ranges that extend beyond Europe, contributing to the goals of other biodiversity sequencing projects. Currently, Lepidoptera are the most extensively sequenced animal order (Figure 1A). Project Psyche is in its first of three phases. Phase 1 aims to generate genomes for 2000 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 1C). 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 5000 species, including whole ecosystem sampling for selected

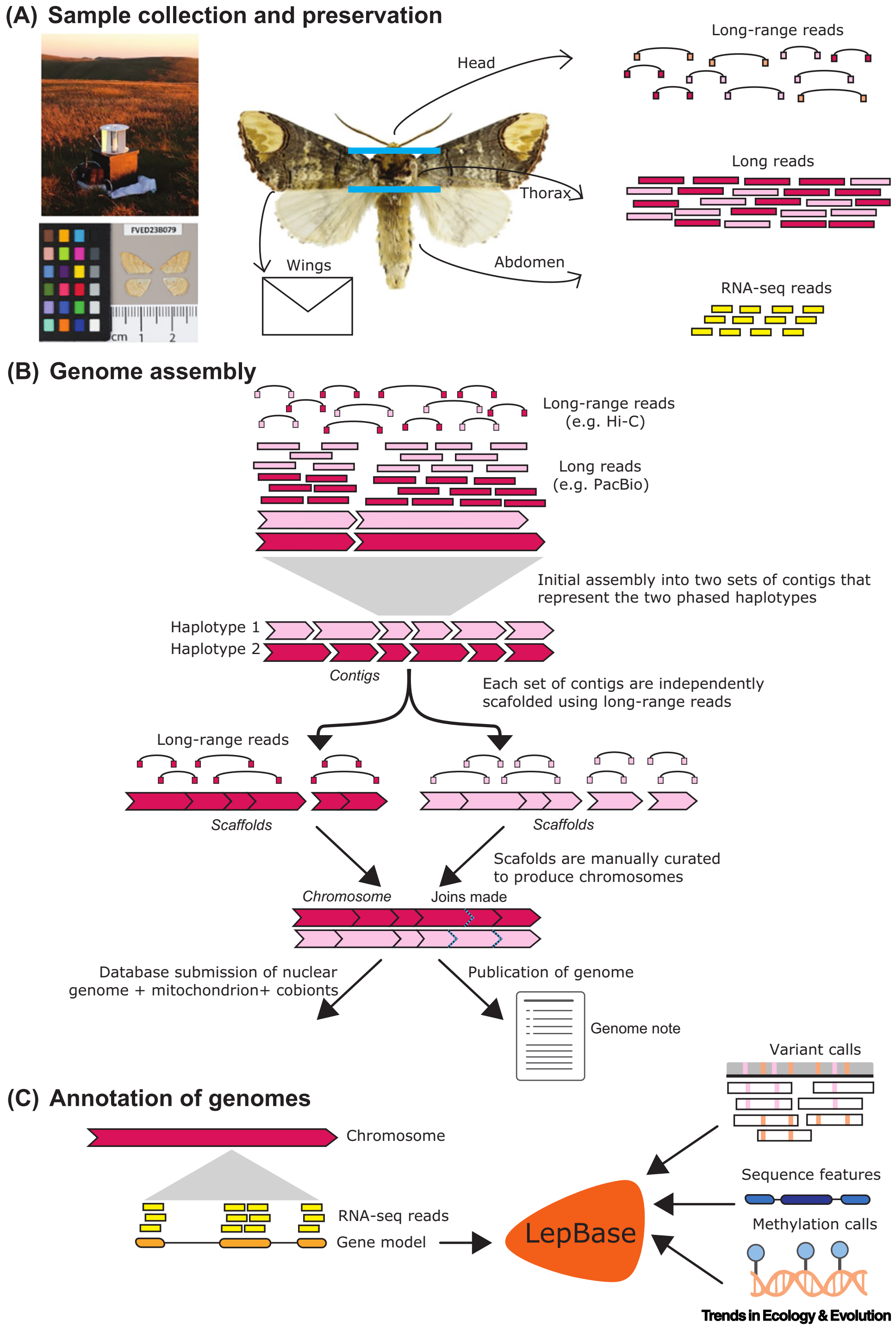


Figure 2. Project Psyche workflow: from catching Lepidoptera to annotated reference genomes. (A) Lepidoptera are caught with traps or nets and photographed in a standardised way. For most specimens the head is used for long-range (Hi-C) sequencing, the thorax for long-read (PacBio) sequencing, and the 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

(Figure legend continued at the bottom of the next page.)

Thomas Decroly: Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, United Kingdom.
Camille Cornet: Biodiversity Genomics Laboratory, University of Neuchâtel, Neuchâtel, Switzerland.
Tinatin Chkhartishvili: Institute of Zoology, Ilia State University, Kakutsa Cholokashvili Ave 3/5, Tbilisi 0179, Georgia.
Richard Challis: Tree of Life Programme, Wellcome Sanger Institute, Cambridge, United Kingdom.
Jon Bridle: Department of Genetics, Evolution and Ecology, University College London, Gower Street, London WC1, United Kingdom.
Bastien Boussau: Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université Lyon 1, UCB Lyon 1 - Bât. Grégor Mendel, 43 bd du 11 novembre 1918, 69622 VILLEURBANNE cedex, France.
Adele Bordoni: Department of Biology, University of Florence, Via Madonna del Piano 6, 50019, Sesto Fiorentino, Firenze, Italy.
Jesper Boman: Evolutionary Biology Program, Department of Ecology and Genetics (IEG), Uppsala University, Evolutionsbiologi, Evolutionsbiologiskt Centrum (EBC), Uppsala universitet, Norbyvägen 18 D, 752 36 Uppsala, Sweden.
Stoyan Beshkov: National Museum of Natural History, Bulgarian Academy of Sciences, Sofia, Bulgaria.
Federica Bertocchini: Plasticentropy, Reims, France.
Saurav Baral: Department of Zoology, Stockholm University, Svante Arrheniusväg 18 B, Stockholm University, S-10691 Stockholm,Sweden.
Joaquín Baixeras: Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Valencia, Spain.
Dominic Absolon: Tree of Life Programme, Wellcome Sanger Institute Cambridge, United Kingdom.

*Corresponding author.
charlotte.wright@sanger.ac.uk (C.J. Wright) and
jm66@sanger.ac.uk (J.I. Meier).

Box 4. Potential alternatives other projects might consider useful

Project Psyche is keen to help other regions establish similar projects and celebrates having sister projects to work synergistically with. Similar initiatives are emerging across the globe, many of which include Lepidoptera within their goals. The EBP unites these projects through a common framework, coordination platforms, and collectively devising strategies to reduce geographical disparities in sequencing and the use of genomic data [140].

Project Psyche is built upon strong existing expertise and infrastructures in Europe, including many salaried scientists studying Lepidoptera and an emerging biodiversity genomics community promoting the use of large-scale data. In less well-resourced regions, it may be helpful to use a targeted approach to generate genomes of key species of economic, biological, conservation, scientific importance, or cultural value. This can draw attention and funding towards larger scale genomics aspirations. While Lepidoptera can typically be sampled, sequenced, and assembled using similar protocols, projects with highly diverse taxonomic groups will require a range of taxon-specific adjustments.

Both widely used long-read sequencing technologies, PacBio and ONT, offer both benchtop and high-throughput sequencing machines. Project Psyche uses PacBio Revio machines for high-throughput long-read sequencing, which massively reduces the sequencing costs compared to lower-throughput machines. However, the high acquisition and maintenance costs of these machines make them most cost-effective if used at capacity. There is thus a benefit to locally centralising sequencing efforts with shared resources. This holds true for short-read sequencing needed for HiC and RNA-seq generation with technologies including Illumina. To reduce costs, genomes can also be annotated without RNA data if well-annotated genomes of related species are available. Long-read sequencers can also be used for long-range sequencing (e.g., Pore-C) if cheaper short-read sequencing options are not available.

In many regions, including tropical biodiversity hotspots, species diversity is much higher than in Europe and depending on the taxonomic group, a large proportion of the species remain undescribed. For certain groups it might be useful to first survey the taxonomic diversity, including potentially cryptic species, and use this information to maximise phylogenetic diversity obtained through reference genome sequencing. One approach for this is a barcode screen, whereby specific genes are amplified with PCR and sequenced for example, with Oxford Nanopore technologies (ONT).

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 ~4000 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 ([33]; www.earthbiogenome.org). Project Psyche aims to develop genomics capacity across Europe and beyond (Box 4). 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.

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, behavioural, phenological, and distributional understanding of Lepidoptera to be integrated for insightful evolutionary inquiry. Crucially, thousands of chromosome-level genomes increase accuracy in detecting and determining genomic changes (e.g., in gene content, transposable elements, or chromosomal rearrangements) associated with phenotypic divergence across 230 million years of evolution. The assemblage of lepidopteran species found in

---

belonging to the same haplotype. Because some regions are not sequenced well, for example, 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 [142]. Image credits: Andy Griffiths (moth trap), Vlad Dinca (standardised photo), Didier Descouens (*Phalera bucephala*).

Glossary

B chromosomes: nonessential 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, including the gut microbiome, symbionts, parasites, and environmental contaminants.

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

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

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.

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

European Widening Countries: EU member states and countries associated with the Horizon Europe Programme that have lower research and innovation performance than others. They are targeted by a set of measures and initiatives in the Programme to reduce these disparities.

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

Heterogametic sex: sex that has two different 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 are the heterogametic sex, typically carrying a Z and a W, whereas males have two Z chromosomes. In some Lepidoptera species, there are multiple Z

Europe constitutes ~7% of all lepidopteran species globally. Nonetheless, with 92/136 families and 27/46 superfamilies represented in Europe (http://www.lepiforum.de), the Project Psyche dataset is an unprecedented resource for addressing evolutionary questions (Figure 1B).

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 $n = 30$–31 chromosomes [37,38], which closely resemble the inferred ancestral linkage groups known as Merian elements [39]. Most species have experienced few chromosome fusions, while fissions are restricted to some groups. However, several lineages have recently undergone extensive genomic rearrangements, with or without dramatic changes in chromosome number [39,40]. 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, their evolutionary constraints [41], and their effects on recombination, adaptation and speciation.

A key question is how holocentricity evolved in Lepidoptera and its sister order Trichoptera. In the silk moth (*Bombyx mori*) a key component of eukaryotic centromeres, CENP-A/CenH3, has been lost [42], while many components of their kinetochores are conserved [43]. A diverse set of genomes will reveal the evolutionary dynamics of centromere and kinetochore composition, including the loss of CENP-A and possible other components. The holocentric architecture requires meiotic adaptations such as restricted kinetochore activity during chromosome segregation [44,45], 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.

In Lepidoptera, males are the homogametic sex with Z chromosomes, while females are heterogametic with ZW chromosomes. Typically, W chromosomes are repeat-rich and historically difficult to assemble. Thanks to high-quality long reads, combined with long-range information and careful manual genome curation, assembling W chromosomes is now routine although obtaining complete W chromosomes is still a challenge (Box 3). Species with multiple degraded sex chromosomes [46] may benefit from cytogenetic confirmation [47]. The Project Psyche data will allow progress on the longstanding debate of how W chromosomes in Lepidoptera arise [48]. 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 [49,50]. Current evidence suggests that W chromosomes evolved multiple times independently, and were lost in some species [49]. Indeed, sex chromosome-autosome fusions are the most frequent type of interchromosomal rearrangement in Lepidoptera [39] and the resulting neo-sex chromosomes can retain homology in regions derived from original autosomes [41,46,51].

As Z chromosomes spend twice as much time in males than in females, they are expected to become enriched in male-biased genes and depleted in female-biased genes due to sexual antagonism [52–54]. Consistent with this, male-biased expression has been observed in neo-Z chromosomes in the Danaini tribe [55,56]. Loci on Z chromosomes are also predicted to evolve more rapidly under positive selection due to their hemizygosity in females [57]. However, the smaller effective population size of Z chromosomes relative to autosomes increases genetic drift, affecting the dynamics of how Z-linked sequences evolve. Faster and more adaptive evolution in the neo-Z chromosomes has been observed in several lepidopteran species [55,58]. As

chromosomes and/or multiple or no W chromosomes.

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.

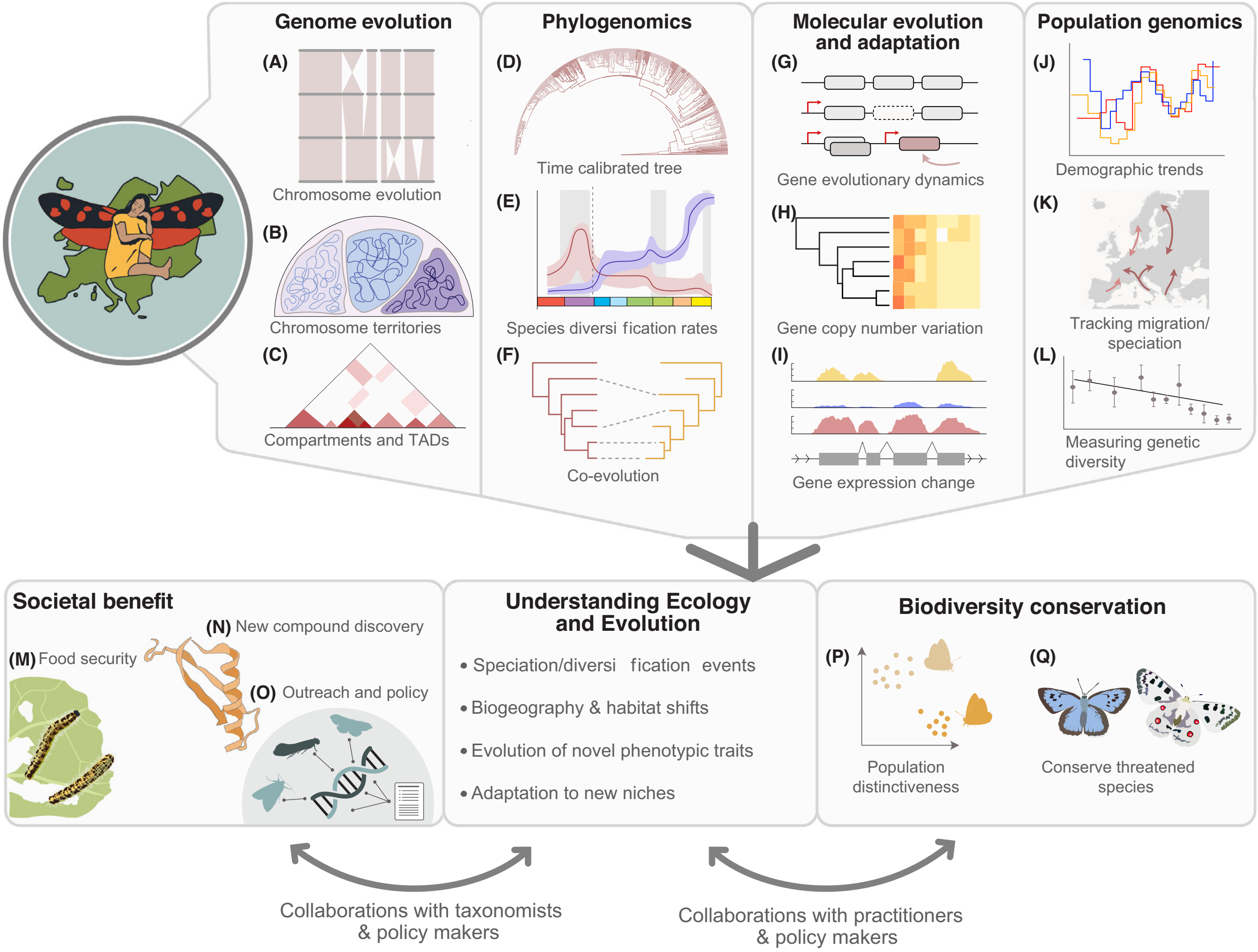
Long-range data: DNA sequence data capturing long-range chromatin interactions within the 3D structure of the nucleus, for example, using Hi-C sequencing. DNA regions that are close to each other in 3D space are crosslinked prior to preparing the DNA for sequencing. Short-read sequencing (e.g., Illumina sequencing) then provides two reads that are from physically close genomic regions. These data are 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.

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

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.



Trends in Ecology & Evolution

Figure 3. 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, and 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 can be a source for human food (especially for protein), either harvested from natural populations, or cultivated. Cultivation and/or domestication has potential for significant expansions driven by new genomic knowledge. In addition, genomics aids tracking and management of lepidopteran 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 reintroductions. (Q) Conservation case studies, large blue *(Phengaris arion)* and apollo butterfly (*Parnassius apollo*).

lepidopteran females likely do not recombine [59], neo-W chromosomes are expected to degenerate over time. The availability of neo-sex chromosomes of various ages in the Project Psyche dataset offers new opportunities to investigate the temporal dynamics of neo-W chromosome degeneration. While pseudogenization, whereby genes become nonfunctional, and epigenetic silencing of nonrecombining W-linked genes may lead to selection for dosage compensation, dosage-sensitive genes can be amplified and moved out of neo-sex chromosomes [60,61]. Population specific degeneration of W-linked genes and associated changes in copy numbers and dosage may promote speciation through the evolution of reproductive incompatibilities [62]. Finally, more than one sex determination mechanism is known in Lepidoptera [63], including a dominant W-linked female determiner in *B. mori* [64], sex determination based on the Z chromosome to autosome ratio in *Samia cynthia* [65], and zygosity-based sex determination in a hypervariable Z locus in *Bicyclus anynana* [66]. W chromosome sequences will shed light on their role in sex determination and the evolution of diverse primary sex-determining mechanisms in Lepidoptera.

The long-range (Hi-C) data used to scaffold Psyche genomes also allows for comparative analyses of genome organisation to include **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. First, **chromosome territoriality** is extremely pronounced with limited interchromosomal contacts (Figure 3B) [67,68]. Second, in addition to the active A and inactive B compartments described in various invertebrates and vertebrates [69], *B. mori* chromosomes also fold into a third compartment, S, with a unique contact pattern [67,68]. The Psyche dataset will allow the investigation of whether strong territoriality and the S compartment are general characteristics of Lepidoptera, and evaluate the 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 so extremely species rich and phenotypically diverse. An exceptional burst of diversification occurred 100–150 million years ago (MYA), where nearly all superfamilies in Ditrysia evolved, comprising 99% of all extant Lepidoptera species [4,70]. This diversification roughly coincides with the origin of flowering plants [2,71], although considerable temporal uncertainty remains. A broad diversity of reference genomes will allow us to infer more accurately dated phylogenies, drawing upon both coding and noncoding regions (Figure 3D,E), 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 [72]. Factors amenable to genomic studies that may facilitate rapid speciation include high genetic variation [73,74], introgression [75], expansion of transposable elements [76], gene birth/death dynamics [77], and chromosomal rearrangements [78], which may cause hybrid dysfunction or recombination suppression [79–81]. 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 [82], but it is unknown if introgression is equally common among slowly speciating lineages. Direct quantification of chromosomal rearrangement rates and comparisons between conserved and rearranged genomes [83], 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, which to date has been elucidated in some selected species. For example, gene duplication enabled Pieridae to overcome the chemical defences of their host plants [84,85] and caused colour polymorphism in the wood tiger moth (*Arctia plantaginis*) [86]. Transposable element insertions led to industrial melanism in the peppered moth (*Biston betularia*) [87] and alternative life-history strategies in the clouded yellow butterfly (*Colias croceus*) [88]. Such 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) [89], whereas introgression facilitated pyrethroid insecticide resistance through gene flow from resistant invading *Helicoverpa armigera* into the native pest species *Helicoverpa zea* in Brazil [90]. 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 complex phenotypic variation, including associations with ants in Lycaenidae, keratin feeding in Tineidae, and many shifts between nocturnal and diurnal activity. 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 [91], 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 complex adaptive phenotypes. Many evolutionary novelties shared among species at the level of genera or deeper, lack relevant variation within species that can be used to gain mechanistic insights via the standard molecular genetics toolkit [92]. The potential of comparative genomics to overcome this challenge is elegantly exemplified by a recent comparative genomic analysis on traits associated with adaptive radiation in Heliconiini butterflies [93]. Analyses of introgression and rates of gene turnover were integrated with scans of selection in coding regions and conserved nonexonic elements, for unprecedented insights into diverse adaptive phenotypes. Project Psyche will allow numerous laboratories to similarly harness the power of comparative genomics, with diverse rigorous insights likely to emerge when coupled with the ease of CRISPR/Cas9 functional validation in Lepidoptera. Thousands of genomes, spanning a large range of divergence times, enable powerful tests of major theoretical predictions including the repeatability of evolution at the genomic and phenotypic level [94,95]. For example, whether a deeper divergence time among taxa experiencing similar selection predicts a lower likelihood of parallelism at the genetic level [96].

We lack a general understanding of how gene regulatory networks are structured and evolve in Lepidoptera, beyond a small number of intensely studied cases, such as the networks involved in wing colour patterning [97] and *Hox* genes [98]. 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 [39,99] but low sequence similarity between noncoding regions [32,100]. Noncoding regions can drive evolutionary novelty by sculpting the transcriptome. For example, a long noncoding RNA generates a miRNA that drives wing colour variation in butterflies [101–103]. Through denser sampling of species, the evolutionary history of noncoding 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 potential impacts on regulatory network outcomes. In addition, chromosome conformation data will enable the investigation of whether regulatory rewiring, through changes in 3D architecture, contributes to trait evolution [104–106] (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 or cause cytoplasmic incompatibilities in their hosts [30,107,108]. The Project Psyche dataset will enable the mapping of broad patterns of cobiont presence and host switching events across Lepidoptera (Figure 3F), building on previous work that indicated limited coevolution between cobionts, such as *Wolbachia*, and their insect hosts [30]. This dataset will also enable investigations into the complex evolutionary dynamics between virus, cobiont, and their hosts. For example, *Wolbachia* itself can be parasitised by double-stranded DNA temperate bacteriophages [109].

Applications to conservation and society

There is an urgent need to understand and predict if, how, and how quickly, populations can respond to a changing environment, and whether ecological communities can persist [110]. 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 genetic underpinnings of evolutionary responses to environmental change that may increase ecological resilience [111]. Psyche genomes also provide a foundation for understanding genetic diversity over time (Figure 3J). A powerful approach will be to combine Project Psyche genomes with resequencing data from many individuals to trace genetic diversity across regions and time, for example, from museum collections (Figure 3K,L). This approach will be used by projects including the European Lepidopteran Population Genomics Consortium (LepEU; https://lepeu.github.io/), which are working to provide insights into population genetic diversity, structure, connectivity, and local adaptation to inform conservation management (Figure 3P) [112]. There is growing consideration of population structure and genetic diversity in addition to taxonomic units in legislation (e.g., the Convention on Biological Diversity), although standardised genetic diversity assessments are yet to be developed [113]. Population genomics will also enable the association of adaptive traits with genomic features, informing captive breeding and genetic rescue efforts [114], as well as the identification of limits to evolutionary rescue [115]. The dense genomic sampling will also permit studies of species delimitation, including within **cryptic species complexes** [116,117], improving our understanding of biodiversity at the continental scale [118] and will 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 accelerate our ability to address societal challenges. This includes biotechnological applications based on Lepidoptera biology. Lepidoptera have already proven to be an invaluable source of potential biotechnological tools in diverse fields including material science and environmental bioremediation. For example, *B. mori* can be engineered to produce highly extensible silk [119]. Several Lepidoptera possess the capacity to degrade plastics [120]. The greater wax worm (*Galleria mellonella*) stands out because of the discovery of enzymes in the animal saliva that can degrade polyethylene, a prevalent and highly resistant type of plastic (Figure 3N) [121–123]. *G. mellonella* is also used as a versatile experimental model for immune responses against human pathogens [124].

As well as critical sources of food security for many human societies [125], with potential for domestication and commercial cultivation in future food economies, Lepidoptera include some of the world's most important pests, causing high economic and societal costs. Reference 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 [90]. Artificial intelligence (AI) can search the genomes for novel targets for highly specific insecticides. However, given the strong impact of insecticides on biodiversity [126] and human health [127], an alternative approach is to implement lower ecological impact strategies including biocontrol and agroecology, benefitting from natural enemies of pests [128]. Likewise, microbiome and virome data [129–131] and genomic insights into how cobionts manipulate host fitness [132,133] are opening new approaches to pest control. Another potential avenue for pest control facilitated by genomes is gene drive [134], although this is not without risks [135]. 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 EBP to sequence all eukaryotic species. With the ever-improving approaches to utilise reference genomes with methods including AI, these reference genomes will provide a rich and vital resource for the ecological and climate challenges of our era.

Outstanding questions

How generalisable are the findings from previous genomic studies based on single species or genera compared with those that encompass genomes from thousands of species?

How and why do lineages differ in species richness, rates of genomic rearrangements, and/or genetic diversity?

What genomic features are associated with major evolutionary transitions, such as shifts between diurnal and nocturnal lifestyles, monophagous or polyphagous feeding behaviours, or aquatic and terrestrial habitats?

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?

What new discoveries in bioengineering will be facilitated by thousands of lepidopteran genomes?

Additional members of the Project Psyche Community

Dominic Absolon, Joaquín Baixeras, Saurav Baral, Federica Bertocchini, Stoyan Beshkov, Jesper Boman, Adele Bordoni, Bastien Boussau, Jon Bridle, Richard Challis, Tinatin Chkhartishvili, Camille Cornet, Thomas Decroly, Laurence Despres, Mónica Doblas-Bajo, Lora Downes, Paula Escuer, Marianne Elias, Brent Emerson, Madeleine Fabusova, Zdenek F. Fric, Jérémy Gauthier, Noémie Hévin, Peter Huemer, Anna K. Hundsdoerfer, Mathieu Joron, Emre Keskin, Eva Kriegova, Evelyn Kustatscher, Sara La Cava, Josep Lancho, Daniel Linke, Lars Littmann, Konrad Lohse, Valéria Marques, Duane McKenna, Mattia Menchetti, Marco Meyer, Tatiana Moreira, Ana Nahirnić-Beshkova, Vazrick Nazari, Rayner Nunez, Antonio S. Ortiz, Erki Õunap, Jaakko Pohjoismäki, Pavel Potocký, Łukasz Przybyłowicz, Mar Repullés, Jadranka Rota, Stefano Scalercio, Benjamin Schattanek-Wiesmar, Petra Schattanek-Wiesmair, Dominik Schmitz, Cibele G. Sotero-Caio, Loukia Spilani, Clara Spilker, Gábor Sramkó, Jenny Stewart, Dejan V. Stojanovic, Alena Suchackova Bartonova, Gerhard Tarmann, Eric Toro-Delgado, Gerhard G. Thallinger, Başak Ulaşlı, Mikko Vallinmäki, Jose Vicente Pérez Santa Rita, Vladimir Visacki, Pavel Vrba, Claudia C. Weber, Martin Wiemers, Marcin Wiorek, Etka Yapar, Milica Zlatkovic, and Giada Zucco.

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.). P.M.-M. and I.K. were supported by the Czech Ministry of Education, Youth and Sports through the sub-programme INTER-COST (grant no. LUC25115). I.A.D receives salary support from the CNRS. P.O.M was supported by a BBSRC Fellowship (UKRI893). C.B. was supported by the Biodiversity Genomics Europe Horizon project. L.D. was supported by funds from the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2 Investment 1.4 project of the National Biodiversity Future Center (NBFC) (code CN_00000033) (spoke 7). R.V. was supported by grant PID2022-139689NB-I00 (MICIU/AEI/10.13039/501100011033 and ERDF, EU) and by grant 2021-SGR-00420 (Departament de Recerca i Universitats, Generalitat de Catalunya). We thank Franz Chandi for help in gathering the data for Figure 1C. 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.

Supplemental information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tree.2025.10.007.

References

1. Van Nieukerken, E.J. *et al.* (2011) Order Lepidoptera Linnaeus, 1758. In *Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness*. *Zootaxa* (3148) (Zhang, Z.-Q., ed.), pp. 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*, Butterfly Conservation Europe
6. Metcalf, R.L., 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, e37502
13. Lewin, H.A. *et al.* (2022) The Earth BioGenome Project 2020: starting the clock. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2115635118
14. Melters, D.P. *et al.* (2012) Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosom. 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.* 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. Hochkirch, A. *et al.* (2022) *European Red List of Insect Taxonomists*, European Union
19. 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.* 114, 367–384
20. Letsch, H. *et al.* (2025) Type genomics: a framework for integrating genomic data into biodiversity and taxonomic research. *Syst. Biol.* Published online May 20, 2025. https://doi.org/10.1093/sysbio/syaf040
21. Garnett, S.T. *et al.* (2018) A spatial overview of the global importance of Indigenous lands for conservation. *Nat. Sustain.* 1, 369–374
22. Mc Cartney, A.M. *et al.* (2022) Balancing openness with Indigenous data sovereignty: an opportunity to leave no one behind in the journey to sequence all of life. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2115860119
23. Mc Cartney, A.M. *et al.* (2023) Indigenous peoples and local communities as partners in the sequencing of global eukaryotic biodiversity. *NPJ Biodivers.* 2, 8
24. Streck, C. (2023) Synergies between the Kunming-Montreal Global Biodiversity Framework and the Paris Agreement: the role of policy milestones, monitoring frameworks and safeguards. *Clim. Pol.* 23, 800–811
25. Houliston, K. *et al.* (2025) Project Psyche sample collection to genome assembly publication workflow. *protocols.io* Published online July 29, 2025. https://www.protocols.io/view/project-psyche-sample-collection-to-genome-assembl-g6cgbzatx
26. Crowley, L. *et al.* (2023) A sampling strategy for genome sequencing the British terrestrial arthropod fauna. *Wellcome Open Res.* 8, 123
27. 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
28. Hartley, M. *et al.* (2022) The BioImage Archive - building a home for life-sciences microscopy data. *J. Mol. Biol.* 434, 167505
29. 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 Res.* 8, 24
30. Vancaester, E. and Blaxter, M. (2023) Phylogenomic analysis of *Wolbachia* genomes from the Darwin Tree of Life biodiversity genomics project. *PLoS Biol.* 21, e3001972
31. Weber, C.C. (2024) Disentangling cobionts and contamination in long-read genomic data using sequence composition. *G3* 14, jkae187
32. Bortoluzzi, C. *et al.* (2023) Lepidoptera genomics based on 88 chromosomal reference sequences informs population genetic

parameters for conservation. *bioRxiv* Published online April 14, 2023. https://doi.org/10.1101/2023.04.14.536868

33. Lawniczak, M.K.N. *et al.* (2022) Standards recommendations for the Earth BioGenome Project. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2115639118
34. Howe, K. *et al.* (2021) Significantly improving the quality of genome assemblies through curation. *Gigascience* 10, giaa153
35. Harrison, P.W. *et al.* (2021) The European Nucleotide Archive in 2020. *Nucleic Acids Res.* 49, D82–D85
36. 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
37. Robinson, R. (1971) *Lepidoptera Genetics*, Pergamon Press
38. Makino, S. (1951) *An Atlas of the Chromosome Numbers in Animals*, Iowa State College Press
39. Wright, C.J. *et al.* (2024) Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nat. Ecol. Evol.* 8, 777–790
40. van der Heijden, E.S.M. *et al.* (2025) Genomics of Neotropical biodiversity indicators: Two butterfly radiations with rampant chromosomal rearrangements and hybridization. *Proc. Natl. Acad. Sci. U.S.A.* 122, e2410939122
41. Wright, C.J. *et al.* (2025) Constraints on chromosome evolution revealed by the 229 chromosome pairs of the Atlas blue butterfly. *Curr. Biol.* 35, 4727–4742.e7
42. Senaratne, A.P. *et al.* (2021) Formation of the CenH3-deficient holocentromere in Lepidoptera avoids active chromatin. *Curr. Biol.* 31, 173–181.e7
43. Cortes-Silva, N. *et al.* (2020) CenH3-independent kinetochore assembly in Lepidoptera requires CCAN, including CENP-T. *Curr. Biol.* 30, 561–572.e10
44. Rosin, L.F. *et al.* (2021) Oligopaint DNA FISH reveals telomere-based meiotic pairing dynamics in the silkworm, *Bombyx mori*. *PLoS Genet.* 17, e1009700
45. Hockens, C. *et al.* (2024) Chromosome segregation during spermatogenesis occurs through a unique center-kinetic mechanism in holocentric moth species. *PLoS Genet.* 20, e1011329
46. Yoshido, A. *et al.* (2020) Evolution of multiple sex-chromosomes associated with dynamic genome reshuffling in Leptidea wood-white butterflies. *Heredity (Edinb).* 125, 138–154
47. Traut, W. *et al.* (2017) Karyotypes versus genomes: the nymphalid butterflies *Melitaea cinxia*, *Danaus plexippus*, and *D. chrysippus*. *Cytogenet. Genome Res.* 153, 46–53
48. 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
49. Han, M.-J. *et al.* (2024) Multiple independent origins of the female W chromosome in moths and butterflies. *Sci. Adv.* 10, eadm9851
50. Dalíková, M. *et al.* (2017) New insights into the evolution of the W chromosome in Lepidoptera. *J. Hered.* 108, 709–719
51. 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.* 41, msae124
52. Rice, W.R. (1987) The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* 41, 911
53. Wright, A.E. *et al.* (2016) How to make a sex chromosome. *Nat. Commun.* 7, 12087
54. Furman, B.L.S. *et al.* (2020) Sex chromosome evolution: so many exceptions to the rules. *Genome Biol. Evol.* 12, 750–763
55. Mongue, A.J. *et al.* (2022) Support for faster and more adaptive Z chromosome evolution in two divergent lepidopteran lineages. *Evolution* 76, 332–345
56. Mora, P. *et al.* (2024) Sex-biased gene content is associated with sex chromosome turnover in Danaini butterflies. *Mol. Ecol.* 33, e17256
57. Charlesworth, B. *et al.* (1987) The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* 130, 113–146
58. Höök, L. *et al.* (2024) Temporal dynamics of faster Neo-Z evolution in butterflies. *Evolution* 78, 1554–1567
59. Turner, J.R. and Sheppard, P.M. (1975) Absence of crossing-over in female butterflies (*Heliconius*). *Heredity (Edinb).* 34, 265–269
60. Potrzebowski, L. *et al.* (2008) Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biol.* 6, e80
61. Síchová, J. *et al.* (2013) Chromosomal evolution in tortricid moths: conserved karyotypes with diverged features. *PLoS One* 8, e64520
62. Filatov, D.A. (2018) The two “rules of speciation” in species with young sex chromosomes. *Mol. Ecol.* 27, 3799–3810
63. Van’t Hof, A.E. *et al.* (2025) Sex determination in moths and butterflies: Masculinizer as key player. *Curr. Opin. Insect Sci.* 70, 101375
64. Kiuchi, T. *et al.* (2014) A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature* 509, 633–636
65. 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
66. Van’t Hof, A.E. *et al.* (2024) Zygosity-based sex determination in a butterfly drives hypervariability of *Masculinizer*. *Sci. Adv.* 10, eadj6979
67. 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
68. Gil, J. *et al.* (2023) Unique territorial and sub-chromosomal organization revealed in the holocentric moth *Bombyx mori*. *bioRxiv* Published online September 16, 2023. https://doi.org/10.1101/2023.09.14.557757
69. Rowley, M.J. and Corces, V.G. (2018) Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800
70. Condamine, F. *et al.* (2016) Global patterns of insect diversification: towards a reconciliation of fossil and molecular evidence? *Sci. Rep.* 6, 19208
71. Mitter, C. *et al.* (2017) Phylogeny and evolution of Lepidoptera. *Annu. Rev. Entomol.* 62, 265–283
72. 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
73. Vellend, M. (2005) Species diversity and genetic diversity: parallel processes and correlated patterns. *Am. Nat.* 166, 199–215
74. 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
75. Peñalba, J.V. *et al.* (2024) The role of hybridization in species formation and persistence. *Cold Spring Harb. Perspect. Biol.* 16, a041445
76. Serrato-Capuchina, A. and Matute, D.R. (2018) The role of transposable elements in speciation. *Genes (Basel)* 9, 254
77. Bista, I. *et al.* (2023) Genomics of cold adaptations in the Antarctic notothenioid fish radiation. *Nat. Commun.* 14, 3412
78. Augustijnen, H. *et al.* (2024) A macroevolutionary role for chromosomal fusion and fission in Erebia butterflies. *Sci. Adv.* 10, eadl0989
79. White, M.J.D. (1978) Modes of speciation. In *A Series of Books in Biology* (456), Freeman, San Francisco
80. Rieseberg, L.H. (2001) Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16, 351–358
81. 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). *Cytogenet. Genome Res.* 153, 213–222
82. Marques, D.A. *et al.* (2019) A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* 34, 531–544
83. Lucek, K. *et al.* (2022) A holocentric twist to chromosomal speciation? *Trends Ecol. Evol.* 37, 655–662
84. 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
85. 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
86. Brien, M.N. *et al.* (2023) Colour polymorphism associated with a gene duplication in male wood tiger moths. *Elife* 12, e80116
87. Van’t Hof, A.E. *et al.* (2016) The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534, 102–105
88. Tunström, K. *et al.* (2023) Evidence for a single, ancient origin of a genus-wide alternative life history strategy. *Sci. Adv.* 9, eabq3713
89. 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

90. 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
91. 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
92. Bomblies, K. and Peichel, C.L. (2022) Genetics of adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2122152119
93. Cicconardi, F. *et al.* (2023) Evolutionary dynamics of genome size and content during the adaptive radiation of Heliconiini butterflies. *Nat. Commun.* 14, 5620
94. Blount, Z.D. *et al.* (2018) Contingency and determinism in evolution: replaying life's tape. *Science* 362, eaam5979
95. Gould, S.J. (1989) *Wonderful Life: the Burgess Shale and the Nature of History*, W.W. Norton and Co.
96. Bohutínská, M. and Peichel, C.L. (2024) Divergence time shapes gene reuse during repeated adaptation. *Trends Ecol. Evol.* 39, 396–407
97. Nadeau, N.J. (2016) Genes controlling mimetic colour pattern variation in butterflies. *Curr. Opin. Insect Sci.* 17, 24–31
98. Mulhair, P.O. *et al.* (2023) Diversity, duplication, and genomic organization of homeobox genes in Lepidoptera. *Genome Res.* 33, 32–44
99. Pringle, E.G. *et al.* (2007) Synteny and chromosome evolution in the lepidoptera: evidence from mapping in *Heliconius melpomene*. *Genetics* 177, 417–426
100. Quah, S. *et al.* (2015) A burst of miRNA innovation in the early evolution of butterflies and moths. *Mol. Biol. Evol.* 32, 1161–1174
101. 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
102. 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
103. Tian, S. *et al.* (2024) A microRNA is the effector gene of a classic evolutionary hotspot locus. *Science* 386, 1135–1141
104. Marlétaz, F. *et al.* (2023) The little skate genome and the evolutionary emergence of wing-like fins. *Nature* 616, 495–503
105. Real, M. *et al.* (2020) The mole genome reveals regulatory rearrangements associated with adaptive intersexuality. *Science* 370, 208–214
106. Lupiáñez, D.G. *et al.* (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012–1025
107. Duplouy, A. and Hornett, E.A. (2018) Uncovering the hidden players in Lepidoptera biology: the heritable microbial endosymbionts. *PeerJ* 6, e4629
108. Jiggins, F.M. *et al.* (2000) The butterfly *Danaus chrysippus* is infected by a male-killing *Spiroplasma* bacterium. *Parasitology* 120, 439–446
109. Gavotte, L. *et al.* (2007) A survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. *Mol. Biol. Evol.* 24, 427–435
110. 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. Ser. B Biol. Sci.* 377, 20210027
111. 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
112. 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
113. Buzan, E. *et al.* (2025) Biodiversity genomics research practices require harmonising to meet stakeholder needs in conservation. *Mol. Ecol.* Published online June 26, 2025. https://doi.org/10.1111/mec.70001
114. Halford, G. *et al.* (2025) Genomic monitoring of a reintroduced butterfly uncovers contrasting founder lineage survival. *Evol. Appl.* 18, e70074
115. 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. Ser. B Biol. Sci.* 380, 20230322
116. Lucek, K. *et al.* (2024) Towards a genomic resolution of the *Phengaris alcon* species complex. *Conserv. Genet.* 25, 823–834
117. 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
118. Vodă, R. *et al.* (2015) Cryptic matters: overlooked species generate most butterfly beta-diversity. *Ecography (Cop.)* 38, 405–409
119. 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
120. Yang, S.-S. *et al.* (2024) Radical innovation breakthroughs of biodegradation of plastics by insects: history, present and future perspectives. *Front. Environ. Sci. Eng.* 18, 1–39
121. Bombelli, P. *et al.* (2017) Polyethylene bio-degradation by caterpillars of the wax moth *Galleria mellonella*. *Curr. Biol.* 27, R292–R293
122. Spínola-Amilibia, M. *et al.* (2023) Plastic degradation by insect hexamerins: near-atomic resolution structures of the polyethylene-degrading proteins from the wax worm saliva. *Sci. Adv.* 9, eadi6813
123. Sanluis-Verdes, A. *et al.* (2022) Wax worm saliva and the enzymes therein are the key to polyethylene degradation by *Galleria mellonella*. *Nat. Commun.* 13, 5568
124. 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
125. Siddiqui, S.A. *et al.* (2023) Edible Lepidoptera as human foods – a comprehensive review. *J. Insects Food Feed* 10, 25–49
126. Rigal, S. *et al.* (2023) Farmland practices are driving bird population decline across Europe. *Proc. Natl. Acad. Sci. U. S. A.* 120, e2216573120
127. 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
128. Goulet, F. *et al.* (2023) Moving beyond pesticides: exploring alternatives for a changing food system. *Environ. Sci. Pol.* 147, 177–187
129. 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
130. Tay, W.T. *et al.* (2023) Spodoptera frugiperda: ecology, evolution, and management options of an invasive species. *Annu. Rev. Entomol.* 68, 299–317
131. Amado, D. *et al.* (2025) The genetic architecture of resistance to flubendiamide insecticide in *Helicoverpa armigera* (Hübner). *PLoS One* 20, e0318154
132. Hoffmann, A.A. *et al.* (2015) Wolbachia strains for disease control: ecological and evolutionary considerations. *Evol. Appl.* 8, 751–768
133. 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
134. Holman, L. (2019) Evolutionary simulations of Z-linked suppression gene drives. *Proc. Biol. Sci.* 286, 20191070
135. Courtier-Orgogozo, V. *et al.* (2020) Evaluating the probability of CRISPR-based gene drive contaminating another species. *Evol. Appl.* 13, 1888–1905
136. Project Psyche Community (2025) *Code of Conduct for members of the Project Psyche Community*, Zenodo
137. 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
138. European Commission: Directorate-General for Research and Innovation (2025) *She figures 2024 – Gender in research and innovation – Statistics and indicators*, Publications Office of the European Union Available from: https://doi.org/10.2777/6847557
139. Laumer, C. (2023) Picogram input multimodal sequencing (PiMmS). Protcols.io Published online January 25, 2023. https://www.protocols.io/view/picogram-input-multimodal-sequencing-pimms-cfaitice
140. Blaxter, M. *et al.* (2025) The Earth BioGenome Project Phase II: illuminating the eukaryotic tree of life. *Front. Sci.* 3, 1514835
141. 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
142. Challis, R.J. *et al.* (2016) Lepbase: the Lepidopteran genome database. *bioRxiv* Published online June 6, 2016. https://doi.org/10.1101/056994