http://ecoevopublisher.com/index.php/ijmec
ISSN 1927-663X

EcoEvo Publisher®

International Journal of Molecular Ecology and Conservation

Vol.16 No.1 2026

2026
01

OPEN ACCESS

Publisher
EcoEvo Publisher

Edited by
Editorial Team of International Journal of Molecular Ecology and Conservation
Email: edit@ijmec.ecoevopublisher.com
Website: http://ecoevopublisher.com/index.php/ijmec

Address:
11388 Stevenston Hwy,
PO Box 96016,
Richmond, V7A 5J5, British Columbia
Canada

International Journal of Molecular Ecology and Conservation (ISSN 1927-663X) is an open access, peer reviewed journal published online by EcoEvoPublisher.

The journal is considering all the latest and outstanding research articles, letters and reviews in all aspects of molecular ecology and conservation, containing the contents of the ranges from the applied to the theoretical in molecular ecology and nature conservation, the policy and management with comprehensive and applicable information; the ecological bases for the conservation of ecosystems, species, genetic diversity, the restoration of ecosystems and habitats; as well as the expands the field of ecology and conservation work.

EcoEvoPublisher is an international Open Access publishing platform that publishes scientific journals in the field of ecology and evolution registered at the publishing platform that is operated by Sophia Publishing Group (SPG), founded in British Columbia of Canada.

International Journal of Molecular Ecology and Conservation (online), 2026, Vol. 16, No.1
ISSN 1927-663X
https://ecoevopublisher.com/index.php/ijmec

Latest Content

International Journal of Molecular Ecology and Conservation, 2026, Vol.16, No.1, 1-12
http://ecoevopublisher.com/index.php/ijmec

Review Article Open Access

Integrative Genomics of Migration, Defense, and Host-Plant Chemistry of the Monarch Butterfly

Jack Sytsma
Kansas State University, Division of Biology, Manhattan KS USA
Corresponding author: sytsmaj134@ksu.edu

International Journal of Molecular Ecology and Conservation, 2026, Vol.16, No.1 doi: 10.5376/ijmec.2026.16.0001
Received: 07 Jan., 2026
Accepted: 04 Feb., 2026
Published: 27 Feb., 2026

Preferred citation for this article:
Sytsma J., 2026, Integrative genomics of migration, defense, and host-plant chemistry of the monarch butterfly, International Journal of Molecular Ecology and Conservation, 16(1): 1-12 (doi: 10.5376/ijmec.2026.16.0001)

Abstract The monarch butterfly (*Danaus plexippus*) represents an unparalleled model for studying the genetic, physiological, and ecological bases of complex adaptive traits. Its multigenerational migration, spanning up to 4,000 km across North America, and its specialized larval dependence on toxic milkweeds (*Asclepias* spp.) exemplify coevolutionary and life-history complexity. Recent advances in genomic and molecular biology have transformed monarch research from natural history to a deeply integrative science. Chromosome-scale genome assemblies, long-read sequencing, and transcriptomic profiling now reveal the genetic architecture underlying migration, diapause, chemical defense, and wing patterning. Functional tools such as RNAi, TALENs, and CRISPR/Cas9 enable causal tests linking candidate genes to behavior and physiology. Population-genomic and selection-scan studies identify polygenic bases for migratory versus resident phenotypes, as well as adaptive divergence related to host-plant chemistry and parasite resistance. Complementary metabolomic analyses elucidate how monarchs sequester, detoxify, and biochemically transform milkweed cardenolides, providing a mechanistic bridge between genotype and ecological function. Emerging integrative frameworks-combining genomics, neurobiology, metabolomics, and ecology-are uncovering how genetic and regulatory networks mediate interactions among monarchs, milkweeds, parasites, and environmental stressors. Future research integrating single-cell neurogenomics, pan-genome analyses, and eco-genomic experiments promises to clarify how these traits evolve and persist amid rapid environmental change. By connecting molecular mechanisms to ecological outcomes, monarch genomics now provides not only a foundation for understanding adaptation and coevolution, but also actionable insights for conserving one of the world’s most iconic migratory insects.

Keywords Conservation genomics; *Danaus plexippus*; Defense chemistry; Metabolomics; Plant-insect interactions

1 Introduction

The monarch butterfly (*Danaus plexippus*) has long captured both scientific and public attention as one of the most charismatic insect species. Its extraordinary, continent-scale migration across North America is unique among butterflies, spanning up to 4,000 km and involves multiple successive generations that complete different stages of the annual cycle (Brower, 1995). This migration is tightly coupled to reproductive diapause and seasonal timing, forming one of the most complex life-history strategies known in insects (Agrawal et al., 2014; 2021; 2025). Equally striking is the monarch’s specialized larval dependence on milkweeds (*Asclepias* spp.), which provide not only nutrition but also chemical protection. Monarch caterpillars ingest toxic cardenolides and sequester them through metamorphosis, rendering both larvae and adults unpalatable to vertebrate predators - a classical example of coevolutionary adaptation and aposematism (Agrawal et al., 2012; 2024; 2025).

For much of the twentieth century, research on monarchs was rooted in natural history, ecology, and physiology - from field studies on migratory navigation (Brower, 1995) and overwintering colonies in Mexico (Rendón-Salinas et al., 2023) to biochemical investigations of cardenolide sequestration (Agrawal et al., 2021; 2024). These ecological and behavioral foundations laid the groundwork for monarchs to emerge as a model system in evolutionary biology (Brower, 1995; Oberhauser et al., 2015). The genomic era began with the publication of the first monarch draft genome (Zhan et al., 2011), which provided a platform for identifying genes and regulatory networks underlying migration, chemical defense, wing patterning, and reproductive physiology. Since then, monarch research has become increasingly integrative, bridging genomics, neuroscience, chemical ecology, and conservation biology.

Advances in genomic and molecular tools have transformed monarch research, enabling detailed investigation of the genetic and regulatory bases of complex traits (Li et al., 2025). High-quality reference genomes, long-read sequencing, and chromosome-scale assemblies now provide near-complete maps of coding and non-coding regions, structural variants, and neo-sex chromosomes (Zhan et al., 2011; Mongue et al., 2017; Zhan et al., 2020). Transcriptomic approaches, including tissue-specific and developmental-stage RNA sequencing, reveal dynamic gene expression patterns underlying migration, diapause, and chemical defense. Functional genomics techniques such as RNA interference, TALENs, and CRISPR/Cas9 genome editing have made it possible to experimentally validate candidate genes and regulatory elements, linking genetic variation to phenotypic outcomes (Markert et al., 2016; Zhang and Reed, 2016). Complementary population genomic analyses, leveraging resequencing, Genome Wide Association Studies (GWAS; Uffelmann et al., 2021), and selection scans, are uncovering polygenic architectures for migratory behavior, adaptation to host-plant chemistry, and responses to parasites. Together, these tools provide a comprehensive framework for connecting genotype to phenotype, enabling integrative studies that span molecular mechanisms, ecological interactions, and conservation applications (Li et al., 2025).

In this review, we synthesize current knowledge of monarch ecology, genetics, and metabolomics (Figure 1) across seven major domains: (1) the development of genomic resources and functional tools, (2) the genetic basis of migration and seasonal behavior, (3) molecular evolution of chemical defense and host-plant specialization, (4) sex-chromosome and structural genome evolution, (5) eco-genomic interactions involving parasites, microbiota, and host plants, (6) population genomic signals that inform conservation, and (7) metabolomic perspectives on host-plant chemistry and monarch sequestration. We also highlight methodological frontiers - from long-read pan-genomics to CRISPR-based functional validation, single-cell genomics, and integrated metabolomics - that are poised to resolve outstanding questions about how the monarch's remarkable life history is encoded in its genome and shaped by ecological interactions. Overall, this review aims not only to synthesize recent advances in monarch genomics and chemical ecology, but also to identify conceptual and methodological gaps that must be addressed to link molecular mechanisms to ecological function and conservation outcomes.

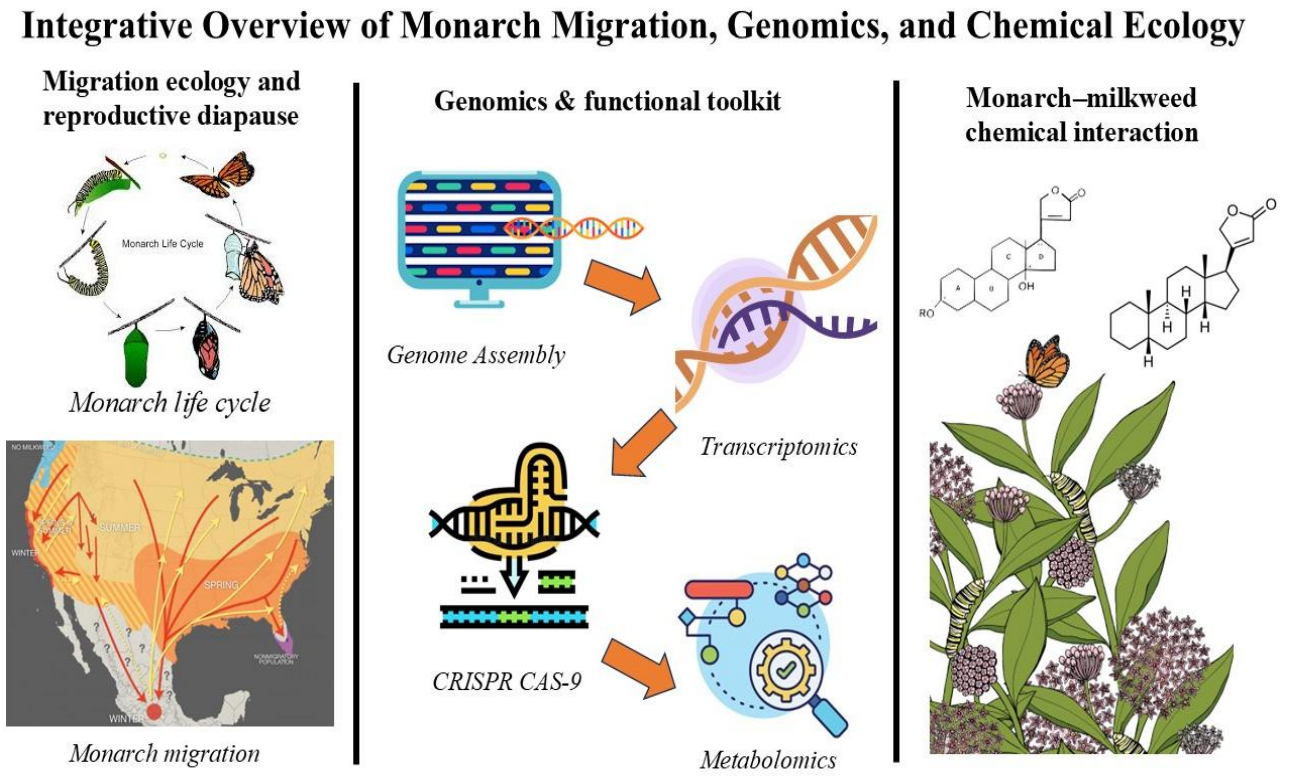

Figure 1 Conceptual model integrating *D. plexippus* migration ecology, genomics, and plant-insect chemical interactions

2 Genomic Resources and Technical Toolkit

2.1 Reference genomes, assemblies, and databases

The initial monarch genome assembly (~273 Mb, ~16 866 predicted protein-coding genes; Table 1) provided a foundational resource for trait mapping and comparative genomics in Lepidoptera (Zhan et al., 2011). Subsequent chromosome-level assemblies, improved annotations, and curated community databases have substantially expanded the utility of monarch genomics for population, functional, and evolutionary analyses (MonarchBase team, 2012; Mongue et al., 2017). These resources have enabled identification of candidate loci underlying migration, circadian regulation, detoxification, pigmentation, and host-plant interactions, as well as comparative analyses of sex chromosome evolution within *Danaus*.

Recent high-quality assemblies have further resolved structural variants, repetitive regions, and the neo-Z chromosome, providing insight into genomic features that may contribute to adaptation and phenotypic divergence among migratory and non-migratory populations (Table 1). Collectively, these monarch-specific genomic resources now support both hypothesis-driven functional studies and broad-scale evolutionary inference.

Table 1 Genomic and functional resources for *D. plexippus* research

Resource type	Description	Year(s)	Applications	Key references
Genome assemblies (draft → chromosome-scale)	Initial draft genome (~273 Mb) followed by improved, chromosome-scale assemblies using long-read sequencing and Hi-C scaffolding; includes annotation of coding genes, repeats, structural variants, and neo-Z chromosome	2011-2020	Trait mapping, comparative genomics, population genomics, sex-chromosome evolution, structural variant discovery	Zhan et al., 2011; Mongue et al., 2017; Zhan et al., 2020
MonarchBase and genomic databases	Community-curated genome browser and annotation resource integrating gene models, transcriptomes, and functional annotations; linked to NCBI and other repositories	2012-present	Gene discovery, annotation refinement, comparative analyses, education and outreach	MonarchBase Team, 2012; Zhan et al., 2011
RNA-seq atlases (tissues, developmental stages)	Bulk RNA-seq from antennae, brain, fat body, flight muscle, wing discs, larvae, pupae, and adults across migratory and reproductive states	2009-present	Gene expression profiling, circadian biology, diapause regulation, migration physiology, developmental genetics	Merlin et al., 2009; Zhan et al., 2011; de Roode et al., 2011
CRISPR/Cas9 and TALEN applications	Targeted genome editing to disrupt or modify candidate genes; functional validation of regulatory and coding loci in monarchs and related Lepidoptera	2016-present	Causal tests of gene function (navigation, pigmentation, circadian clocks), regulatory element validation	Markert et al., 2016; Zhang and Reed, 2016
Metabolomic datasets (milkweeds and monarchs)	LC-MS/MS and untargeted metabolomics of milkweed secondary metabolites and monarch tissues; quantification of cardenolide diversity, sequestration, and biotransformation	2013-present	Chemical ecology, host-plant adaptation, parasite resistance, eco-genomic integration	Petschenka et al., 2013; Dreisbach et al., 2023; Agrawal et al., 2025
OE parasite genomic resources	Genomic and transcriptomic resources for *Ophryocystis elektroscirrha*, a specialist protozoan parasite of monarchs	2015-present	Host-parasite coevolution, disease ecology, immunity and chemical defense interactions	de Roode et al., 2008; Satterfield et al., 2015

2.2 Functional genomics, genome editing, and multi-omic tools

Functional genomic studies in monarchs have leveraged transcriptomic, proteomic, and metabolomic data to link genetic variation with key migratory, physiological, and defensive traits (Table 1). For example, tissue-specific transcriptomic analyses of antennae revealed circadian clock gene expression patterns essential for time-compensated sun-compass navigation (Merlin et al., 2009), while expression profiling of fat body and flight muscle tissues clarified metabolic shifts associated with long-distance migration and lipid utilization (de Roode et al., 2011).

Monarch research has also been at the forefront of integrating chemical ecology with genomics. Metabolomic analyses of larvae and adults have quantified cardenolide sequestration and detoxification, revealing how host-plant chemistry interacts with monarch genotype to shape toxin resistance and performance (Agrawal et al., 2012; Petschenka et al., 2013; Dreisbach et al., 2023). Population genomic resequencing across migratory and resident populations has identified loci associated with neural, metabolic, and endocrine function, providing candidates for functional validation (Zhan et al., 2014).

Genome editing approaches, including TALENs and CRISPR/Cas9, have enabled direct tests of gene function in monarchs and related Lepidoptera, allowing researchers to move from association-based inference toward causal understanding of traits such as circadian regulation and pigmentation (Markert et al., 2016; Zhang and Reed, 2016). Together, these monarch-focused functional and multi-omic applications illustrate how established technologies can be combined to address uniquely integrative questions spanning behavior, chemistry, and ecology.

3 Genetic Architecture of Migration and Seasonal Behavior

3.1 Migratory phenotype complexity

Monarch migration is a compound phenotype encompassing sun-compass orientation, reproductive diapause, lipid storage, flight endurance, and seasonal timing (Table 2). Sun-compass orientation depends on antennal circadian clocks and neural integration of solar cues, as demonstrated by experimental disruption of antennal function or clock genes (Merlin et al., 2009; 2020). Reproductive diapause is regulated by photoperiod and hormone signaling (Fleming and Alto, 2006; Green et al., 2019; Freedman et al., 2023), while lipid storage and flight endurance are shaped by metabolic and mitochondrial pathways associated with long-distance flight capacity (de Roode et al., 2011; Zhan et al., 2014). Seasonal timing of migration and breeding integrates environmental cues such as day length and temperature across generations. Collectively, these traits arise from interacting circadian, endocrine, metabolic, neural, and developmental modules, reflecting a modular and polygenic architecture underlying the monarch's migratory phenotype.

3.2 Antennal clocks, sun-compass orientation, and molecular candidates

Peripheral circadian clocks located in the antennae are essential for time-compensated sun-compass navigation, allowing monarchs to adjust orientation as the solar position changes throughout the day (Merlin et al., 2009; Hemstrom et al., 2025). These clocks interact with central brain regions involved in spatial and sensory integration, including the central complex and optic lobes (Merlin et al., 2009; Guerra et al., 2012). Core clock genes (e.g., *period*, *timeless*, *cryptochromes*) coordinate rhythmic gene expression underlying behavioral timing, while transcriptomic analyses implicate broader networks of sensory and metabolic genes contributing to energy allocation and flight performance (Zhan et al., 2011; Zhan and Reppert, 2013). Together, these findings highlight migration as an emergent property of interconnected circadian, sensory, and metabolic pathways rather than a single master regulator.

3.3 Population genomics: migratory vs. resident populations

Comparative genomic analyses of migratory and resident monarch populations reveal modest but consistent differentiation at loci associated with neural signaling, lipid metabolism, and endocrine function, consistent with adaptation to migratory versus sedentary lifestyles (Zhan et al., 2014; de Roode et al., 2013). Divergence at endocrine-related loci involved in juvenile hormone signaling may contribute to differences in reproductive diapause between populations (Hemstrom et al., 2025). Most signals are dispersed across the genome, supporting a largely polygenic architecture for migration, although a small number of loci with larger effect sizes may disproportionately influence key migratory traits (Freedman and Kronforst, 2023).

3.4 Toward causal tests

Advances in genomic mapping, selection experiments, and functional genetic approaches provide promising avenues for testing causal links between genotype and migratory phenotypes. Genome-wide association studies, artificial selection lines, and pedigreed crosses can be combined with CRISPR/Cas9-based validation to interrogate candidate genes involved in circadian regulation, sensory processing, and energy metabolism (Di

Cristina et al., 2025). Successful genome editing in monarchs and related Lepidoptera demonstrates the feasibility of moving from genotype-phenotype correlations toward mechanistic understanding of migratory and sensory traits (Markert et al., 2016; Zhang and Reed, 2016).

Table 2 Major genes and molecular pathways leading to diverse *D. plexippus* phenotypes

Category	Resource, gene, or pathway	Trait or biological process	Evidence type	Key references
Genomic resources	Draft and chromosome-scale genome assemblies	Genome organization, migration, chemical defense, sex-chromosome evolution	Comparative genomics, population genomics	Zhan et al., 2011; Mongue et al., 2017; Zhan et al., 2020
Databases	MonarchBase and associated repositories	Gene annotation, transcriptome access	Genome curation, comparative analysis	MonarchBase Team, 2012
Functional genomics	RNA-seq atlases (antennae, brain, fat body, wings)	Circadian rhythms, diapause, flight metabolism	Differential expression	Merlin et al., 2009; de Roode et al., 2011
Genome editing	CRISPR/Cas9 and TALENs	Causal testing of candidate genes	Knockout, allele disruption	Markert et al., 2016; Zhang and Reed, 2016
Migration (circadian clock)	period (per), timeless (tim), cryptochrome 2 (cry2)	Sun-compass orientation, migratory timing	Expression, functional assays	Merlin et al., 2009; Guerra et al., 2012
Migration (sensory integration)	Orco and sensory receptor pathways	Orientation and navigation	Expression, candidate gene inference	Zhan et al., 2011; Zhan et al., 2014
Migration (metabolism and diapause)	Insulin signaling (IGF2), juvenile hormone pathways	Lipid storage, reproductive diapause	GWAS, expression, hormone manipulation	Zhan et al., 2014; Freedman and Kronforst, 2023
Chemical defense	Na^+/K^+-ATPase (ATPα) substitutions	Cardenolide resistance	Biochemical assays, comparative genomics	Petschenka et al., 2013; Agrawal et al., 2012
Detoxification and transport	Cytochrome P450s, ABC transporters	Sequestration and biotransformation of toxins	Expression, metabolomics	Petschenka and Agrawal, 2015; Dreisbach et al., 2023
Metabolomics	Milkweed and monarch metabolite profiles	Host-plant adaptation, parasite resistance	LC-MS/MS, untargeted metabolomics	Dreisbach et al., 2023; Agrawal et al., 2025
Sex chromosomes	Neo-Z chromosome, doublesex and hormone signaling genes	Sex-biased expression, genome evolution	Long-read genomics, expression	Mongue et al., 2017
Eco-genomic interactions	Immunity genes, OE parasite, microbiome pathways	Parasite resistance, fitness trade-offs	Infection assays, RNA-seq, metabolomics	de Roode et al., 2008; Hammer et al., 2014
Conservation genomics	Adaptive alleles and metabolite indicators	Population resilience, migration persistence	Population genomics, metabolomic monitoring	Semmens et al., 2016; Thogmartin et al., 2017

4 Chemical Defense: Na+/K+-ATPase Evolution and Cardenolides

Monarch larvae sequester cardenolides from milkweeds, which bind and inhibit Na^+/K^+-ATPase (ATPα; Mongue et al., 2025). Specific amino-acid substitutions in ATPα, such as *N122H* and *Q111L*, reduce binding affinity for cardenolides and confer resistance (Petschenka et al., 2013; López-Goldar et al., 2024). Convergent evolution has been observed in other specialist herbivores (Agrawal et al., 2024), like *Danaus chrysippus* and *Tetraopes* beetles, which carry similar substitutions conferring toxin resistance. Biochemical assays have demonstrated that these substitutions maintain ion pump function while reducing cardenolide binding, illustrating a clear genotype-phenotype link. Variation in ATPα selectivity among monarch populations correlates with milkweed species in their breeding ranges, highlighting an eco-genomic interaction between host-plant chemistry and monarch defense strategies (Petschenka et al., 2013; Agrawal et al., 2012; 2024).

Metabolomic studies of milkweed species reveal extensive variation in cardenolide composition and other secondary metabolites, which can interact with monarch genotype to influence sequestration efficiency, developmental success, and predator deterrence (Malcolm and Brower, 1989; Petschenka and Agrawal, 2015; Agrawal et al., 2024). Integrating chemical and genomic data allows a detailed understanding of coevolution between monarchs and their host plants.

5 Pigmentation, Mimicry, and Developmental Genetics

Monarch wing coloration exemplifies aposematism, with a consistent orange-black pattern across populations that reinforces predator avoidance in conjunction with cardenolide-based chemical defense. Although pigmentation genetics is not known to contribute directly to migratory behavior, it plays an important indirect role in defense by enhancing the effectiveness of warning coloration and mimicry systems. Functional studies suggest that genes such as *optix*, *cortex*, and *WntA*, known from *Heliconius* butterflies, regulate pigment deposition and patterning in monarchs (Martin et al., 2012; Livraghi et al., 2024; 2025). Regulatory changes in cis-elements of these loci likely modulate spatial expression of pigment genes rather than altering protein sequences directly, a pattern observed broadly across Lepidoptera (Ben Chehida et al., 2025). Transcriptomic analyses of wing discs have identified differential expression of *yellow* and *dopa decarboxylase* during late pupal stages, linking enzymatic pathways to melanin and ommochrome deposition (Shen, 2024). Comparative studies across *Danaus* species may further elucidate the molecular basis of subtle color pattern variation relevant to predator learning and mimicry (De-Kayne et al., 2025).

6 Sex Chromosomes and Structural Genome Evolution

Monarchs possess a neo-sex chromosome (neo-Z) resulting from an autosome-Z fusion approximately 5-10 million years ago (Mongue et al., 2017; Mora et al., 2024). Genes on the neo-Z show accelerated evolution and increased sex-biased expression compared to autosomes. For example, *doublesex* and genes involved in hormone signaling are enriched on the neo-Z (Mora et al., 2024), potentially influencing sexual dimorphism in wing coloration and reproductive timing. Long-read sequencing has identified inversions and structural variants on the neo-Z that may reduce recombination and preserve co-adapted gene complexes related to migration and diapause. Transposable element (TE) accumulation on sex chromosomes also contributes to structural diversification, shaping the evolutionary trajectory of *Danaus* genomes (Davey et al., 2016).

7 Parasites, Microbiome, and Eco-genomic Interactions

The specialist protozoan *Ophryocystis elektroscirrha* (OE) reduces monarch survival, flight performance, and fecundity (Altizer and Oberhauser, 1999; Agrawal et al., 2012). Studies show that monarchs feeding on high-cardenolide milkweeds carry lower parasite loads, linking chemical defense to disease resistance (de Roode et al., 2008; Müller-Theissen et al., 2025). Metabolomic analyses of milkweed secondary compounds, including digitoxin, asclepin, and calotropin, reveal variation in toxicity and sequestration efficiency across populations. Monarch gut microbiomes, dominated by *Enterococcus* and *Lactobacillus* species, influence digestion and detoxification pathways (Sanaei et al., 2024); RNA-seq profiling suggests microbial modulation of host immunity and nutrient assimilation (Hammer et al., 2014; van der Hoeven et al., 2013). Integration of host, parasite, plant chemistry, and microbiome data provides a holistic view of eco-genomic dynamics.

8 Population Genomics, Demography, and Conservation Relevance

Large-scale population genomic analyses indicate that eastern North American monarchs exhibit modest population structure, with subtle differentiation between migratory and non-migratory populations (Zhan et al., 2014; Freedman and Kronforst, 2023). Loci associated with lipid metabolism, circadian regulation, and neural function show signals consistent with local adaptation linked to migratory behavior. Demographic modeling further suggests that overwintering populations in Mexico have declined by more than 80% over the past two decades, underscoring ongoing conservation concern (Semmens et al., 2016; Thogmartin et al., 2017). Genomic diversity metrics are increasingly used to identify conservation-relevant units and track the retention of adaptive variation related to migration and reproduction. Integrating these genomic insights into management strategies-such as habitat restoration and protection of migratory corridors-remains an important but unresolved challenge (U.S. Fish and Wildlife Service, 2020; Erickson et al., 2023).

9 Outstanding Gaps and Priority Directions

Despite remarkable advances in monarch genomics and functional biology, several critical gaps remain that limit our mechanistic understanding of migration, chemical defense, and adaptation to environmental stressors.

9.1 Causal mapping of migratory behavior

Migration in monarchs is increasingly recognized as a polygenic and modular trait shaped by circadian regulation, neural circuitry, endocrine signaling, and metabolic pathways (Table 2). Although population genomic and quantitative genetic studies have identified candidate loci associated with migratory components, direct tests of causality remain limited (Freedman and Kronforst, 2023). Integrating high-resolution mapping with functional validation-such as allele-specific perturbations targeting genes implicated in orientation, diapause, lipid storage, and flight endurance-will be essential for resolving how combinations of alleles produce coordinated migratory phenotypes (Markert et al., 2016).

9.2 Regulatory and cell-type-specific neurogenomics

Migratory orientation and seasonal behavior are mediated by complex neural circuits, particularly in the antennae and central brain. Single-cell RNA sequencing (scRNA-seq), spatial transcriptomics, and ATAC-seq of neural tissues under different photoperiods and temperatures can reveal cell-type-specific regulatory programs. Integration with functional assays will clarify how environmental cues are encoded at the molecular level to influence migration timing and sun-compass navigation (Merlin et al., 2009; Guerra et al., 2012; Agrawal et al,. 2024).

9.3 Pan-genome and structural variation analyses

Recent discoveries of chromosomal rearrangements, including a neo-Z chromosome, highlight the potential importance of structural variation in monarch adaptation (Höök et al., 2024). However, most genomic inferences still rely on a single reference genome. Pan-genome approaches incorporating migratory and non-migratory populations will enable systematic assessment of structural polymorphisms, copy-number variation, and transposable elements that may influence migration, diapause, or chemical defense (Davey et al., 2016; Mongue et al., 2017). A key unresolved question is how such variants contribute to adaptation through regulatory versus coding effects.

9.4 Host-microbe-parasite-plant interactions

Monarch fitness and migration are tightly linked to interactions with milkweed chemistry, gut microbiota, and the protozoan parasite *Ophryocystis elektroscirrha* (OE; Müller-Theissen et al., 2025). Controlled factorial experiments combining milkweed species with differing cardenolide profiles, manipulations of the microbiome, and OE exposure can dissect the genetic and metabolic bases of host tolerance, detoxification, and parasite resistance (de Roode et al., 2008; Satterfield et al., 2015; Müller-Theissen et al., 2025).

9.5 Metabolomics and chemical ecology of host plants and monarchs

A particularly understudied frontier is integrating metabolomics to link plant chemistry**,** monarch metabolism, and stress responses. Milkweeds produce a diverse array of cardenolides, alkaloids, and other secondary metabolites (Hoogshagen et al., 2024) that vary among species, populations, and environmental conditions. Monarch larvae ingest and sequester these compounds (Agrawal et al., 2012; Betz et al., 2025), but the dynamics of sequestration, biotransformation, and excretion remain poorly characterized. Mass spectrometry-based metabolomics, including LC-MS/MS and untargeted metabolite profiling, can quantify both plant and insect chemical landscapes, revealing:

How monarchs metabolically adjust to high vs. low cardenolide diets.

The impact of environmental stressors (temperature, drought, pesticide exposure) on metabolite accumulation and detoxification pathways.

Relationships between metabolite profiles and parasite resistance, flight performance, or survival during overwintering.

Combining metabolomics with transcriptomics and proteomics can map the regulatory networks linking detoxification enzymes (e.g., *cytochrome P450s, ABC transporters)* to physiological and behavioral traits, providing a mechanistic understanding of how monarchs cope with chemically complex host plants.

9.6 Conservation integration

Translating molecular, genomic, and metabolomic insights into actionable conservation strategies remains a critical priority for monarch research (Box 1). Although habitat loss and climate change are widely recognized drivers of monarch declines, increasing evidence suggests that the *quality*, chemical composition, and genetic compatibility of restored habitats may be as important as habitat quantity (Thogmartin et al., 2017; Erickson et al., 2023). Milkweed species and genotypes vary substantially in cardenolide concentration, chemical composition, and inducibility under environmental stress, with downstream effects on monarch detoxification, growth, and resistance to the protozoan parasite *Ophryocystis elektroscirrha* (OE; Agrawal et al., 2012; de Roode et al., 2008; Petschenka and Agrawal, 2015). Metabolomic profiling of candidate milkweeds used in restoration programs could therefore identify plant chemotypes that balance larval performance, chemical defense, and disease suppression, while avoiding unintended consequences such as selecting for excessively toxic plants that impair development or migratory performance (Malcolm and Brower, 1989; Agrawal et al., 2024; 2025).

At the population level, coupling genomic and metabolomic data across landscapes offers a complementary framework for conservation planning. Population genomic analyses reveal consistent signals of selection on loci associated with migration, lipid metabolism, circadian rhythms, and endocrine regulation (Zhan et al., 2014; Freedman and Kronforst, 2023), suggesting that adaptive capacity may erode before demographic declines become evident. Integrating these genomic markers with metabolomic signatures-such as cardenolide sequestration profiles, lipid reserves, and stress-response metabolites-could help identify populations that retain key functional traits required for long-distance migration and overwintering success, while providing early warning indicators of declining resilience (Semmens et al., 2016; Thogmartin et al., 2017). Together, advances in causal genomics, pan-genomics, regulatory neurogenomics, and metabolomics position monarchs as a powerful model for linking genotype, chemistry, behavior, and fitness across ecological scales, informing conservation strategies that preserve both population size and the functional diversity required to sustain migration under ongoing environmental change (U.S. Fish and Wildlife Service, 2020; Erickson et al., 2023).

10 Recommended Approaches and Experimental Roadmap

To advance understanding of monarch migration, chemical defense, and adaptation, research should integrate high-resolution genomic, functional, and ecological approaches. Large, replicated mapping cohorts, including pedigreed crosses and population-resequencing panels, can identify loci underlying complex traits such as migratory orientation, diapause, and lipid storage. Functional validation via CRISPR/Cas9 or TALEN-mediated knockouts and allele swaps will allow causal testing of candidate genes. Complementary single-cell and spatial transcriptomics of antennae and brain tissues can reveal cell-type-specific regulatory programs controlling navigation, circadian timing, and sensory processing.

At the genomic level, long-read sequencing combined with Hi-C scaffolding enables chromosome-scale assemblies and pan-genome analyses, facilitating discovery of structural variants (Livraghi et al., 2024), neo-sex chromosome polymorphisms (Mongue et al., 2017), and transposable element dynamics (De-Kayne et al., 2025). SV-aware association testing can link structural changes to adaptive traits. Integrating these genomic data with metabolomic and proteomic profiling will illuminate the biochemical pathways through which monarchs process milkweed toxins (Agrawal et al., 2012; 2024; 2025) and respond to environmental stressors (Dalla et al., 2014), including parasite exposure (Altizer et al., 2015) and habitat change (Green and Kronforst, 2019).

Finally, experimental designs should bridge molecular and ecological scales. Field and laboratory studies combining genotype, microbiome composition, host-plant chemistry (Dale and Stumpe, 2014), and parasite load can quantify the ecological relevance of genetic and metabolic variation. This integrative roadmap positions future research to resolve mechanistic links from genotype to phenotype to fitness, while providing actionable insights for conservation management, such as selecting milkweed species and populations that optimize monarch survival, detoxification capacity, and migratory performance.

11 Conclusions

Research on the monarch butterfly has progressed from classical natural history and ecological observations to an integrative genomic, functional, and eco-physiological understanding of this iconic species. Chromosome-scale genome assemblies, population resequencing, and functional genomic tools have enabled identification of genes and pathways associated with migratory behavior, circadian rhythms, chemical defense, and sex-chromosome evolution (Satterfield et al., 2015). Studies of host-plant interactions and metabolomics have revealed how cardenolide sequestration and chemical stress responses mediate survival, predator avoidance, and parasite resistance, highlighting the complex interplay between genotype, phenotype, and environment (Agrawal et al., 2012; 2024; 2025). Simultaneously, ecological and population genomic analyses underscore how microbiomes, parasite pressures, and environmental changes shape adaptive variation and influence conservation priorities (de Roode et al., 2008; Dale et al., 2014; Sanaei et al., 2024).

Despite these advances, significant gaps remain in fully elucidating the proximate and ultimate mechanisms underlying monarch adaptation. Key opportunities include causal mapping of migration-related alleles, single-cell and spatial genomics to resolve cell-type specific regulatory networks, pan-genome analyses of structural variants and neo-sex chromosome evolution. Additional metabolomic profiling to link host-plant chemistry with physiological stress responses. Integrative experimental designs combining genotype, metabolome, microbiome, milkweed chemistry, and parasite exposure (Dreisbach et al., 2023; Agrawal et al., 2025) offers a pathway to connect molecular mechanisms with ecological function. By leveraging these multi-dimensional approaches, future research can provide mechanistic insights into migration, chemical defense, and adaptation while directly informing conservation strategies to enhance monarch resilience in the face of habitat loss, climate change, and shifting ecological pressures.

Acknowledgments

I thank colleagues and mentors at Kansas State University and Central College for insightful discussions that helped shape the ideas synthesized in this review. I am also grateful to the anonymous reviewer for their thoughtful and constructive feedback, which improved the clarity and scope of the manuscript. I thank the researchers in the monarch biology, genomics, and chemical ecology communities whose foundational and ongoing work made this synthesis possible. No specific funding sources directly supported the preparation of this manuscript.

Conflict of Interest Disclosure

The author declares no conflicts of interest.

References

Agrawal A.A., Petschenka G., Bingham R.A., Weber M.G., and Rasmann S., 2012, Toxic cardenolides: chemical ecology and coevolution of specialized plant-herbivore interactions, Annual Review of Ecology, Evolution, and Systematics, 43: 59-80.

Agrawal A.A., Böröczky K., Haribal M., Hastings A.P., White R.A., Jiang R.W., and Duplais C., 2021, Cardenolides, toxicity, and the costs of sequestration in the coevolutionary interaction between monarchs and milkweeds, Proceedings of the National Academy of Sciences, 118(16): e2024463118.
https://doi.org/10.1073/pnas.2024463118

Agrawal A.A., Hastings A.P., and Duplais C., 2024, Testing the selective sequestration hypothesis: Monarch butterflies preferentially sequester plant defences that are less toxic to themselves while maintaining potency to others, Ecology Letters, 27(1): e14340.
https://doi.org/10.1111/ele.14340

Agrawal A.A., Hastings A.P., Lenhart P.A., Blecher M., Duplais C., Petschenka G., Hawlena D., Wagschal V., and Dobler S., 2024, Convergence and divergence among herbivorous insects specialized on toxic plants: revealing syndromes among the cardenolide feeders across the insect tree of life, The American Naturalist, 204(3): 201-220.
https://doi.org/10.1086/731277

Agrawal A.A., Hastings A.P., Patrick E.T., Knight A.C., and Dreisbach T.A., 2025, Metabolomic variation in milkweeds shapes herbivore defense, performance, and parasite resistance, Ecology Letters.

Agrawal A.A., Salminen J.P., and Fishbein M., 2012, The coevolution of plant defenses and insect herbivores, Ecology, 93(3): 1-10.

Altizer S., and Oberhauser K., 1999, Effects of the protozoan parasite Ophryocystis elektroscirrha on the survival and reproduction of monarch butterflies, Journal of Invertebrate Pathology, 74(1): 61-68.
https://doi.org/10.1006/jipa.1999.4853

Altizer S., Hobson K.A., Davis A.K., De Roode J.C., and Wassenaar L.I., 2015, Do healthy monarchs migrate farther? Tracking natal origins of parasitized vs. uninfected monarch butterflies overwintering in Mexico, PLOS ONE, 10(11): e0141371.
https://doi.org/10.1371/journal.pone.0141371

Ben Chehida Y., van der Heijden E.S., Page E.J., Salazar C.P.A., Rosser N., Gabriela Gavilanes Córdova K., et al., 2025, Genetic parallelism underpins convergent mimicry coloration across Lepidoptera, bioRxiv, 2025-06.
https://doi.org/10.1101/2025.06.26.661542

Betz A., Höglinger B., Walker F., and Petschenka G., 2025, Regionality and temporal dynamics of sequestration and relocation of cardenolides in the monarch butterfly, *Danaus plexippus*, Journal of Chemical Ecology, 51(1): 19.
https://doi.org/10.1007/s10886-025-01572-8

Brower L.P., 1995, Understanding and misunderstanding the migration of the monarch butterfly (Nymphalidae) in North America: 1857-1995, Journal of the Lepidopterists' Society, 49: 304-385.

Dale J., and Stumpe M., 2014, Host plant toxicity, predator-prey interactions, and the evolution of defense: a synopsis, Ecology Letters, 17(7): 831-842.

Dalla S., and Edger P.P., 2014, Evolution of the plant Na+/K+-ATPase gene family: evidence for functional diversification, BMC Genomics, 15: 1-13.

Davey J.W., Chouteau M., Barker S.L., Maroja L., Baxter S.W., Simpson F., Merrill R.M., Joron M., Mallet J., and Dasmahapatra K.K., et al., 2016, Major improvements to the Heliconius melpomene genome assembly reveal multiple chromosomal fusions in Lepidoptera, G3: Genes, Genomes, Genetics, 6(3): 695-708.
https://doi.org/10.1534/g3.115.023655

de Roode J.C., Pedersen A.B., Hunter M.D., and Altizer S., 2008, Host plant species affects virulence in monarch butterfly parasites, Journal of Animal Ecology, 77(2): 120-127.
https://doi.org/10.1111/j.1365-2656.2007.01305.x

De-Kayne R., Gordon I.J., Terblanche R.F., Collins S., Saitoti Omufwoko K., Martins D.J., and Martin S.H., 2025, Incomplete recombination suppression fuels extensive haplotype diversity in a butterfly colour pattern supergene, PLoS Biology, 23(2): e3003043.
https://doi.org/10.1371/journal.pbio.3003043

Di Cristina G., Dirksen E., Altenhein B., Büschges A., and Korsching S.I., 2025, Pioneering genome editing in parthenogenetic stick insects: CRISPR/Cas9-mediated gene knockout in Medauroidea extradentata, Scientific Reports, 15(1): 2584.
https://doi.org/10.1038/s41598-025-85911-5

Dreisbach D., Bhandari D.R., Betz A., Tenbusch L., Vilcinskas A., Spengler B., and Petschenka G., 2023, Spatial metabolomics reveal divergent cardenolide processing in the monarch (*Danaus plexippus*) and the common crow butterfly (Euploea core), Molecular Ecology Resources, 23(6): 1195-1210.
https://doi.org/10.1111/1755-0998.13786

Dreisbach T.A., Hastings A.P., and Agrawal A.A., 2023, Untargeted metabolomics reveals chemical diversity and inducibility of cardenolides across milkweed species, Journal of Chemical Ecology, 49(4): 345-360.

Erickson E., Jason C., Machiorlete H., de la Espriella L., Crone E.E., and Schultz C.B., 2023, Using community science to map western monarch butterflies (*Danaus plexippus*) in spring, Ecology and Evolution, 13(12): e10766.
https://doi.org/10.1002/ece3.10766

Freedman M.G., and Kronforst M.R., 2023, Genomic insights into the evolution and loss of migration in monarch butterflies, Molecular Ecology, 32(8): 1950-1966.

Freedman M.G., and Kronforst M.R., 2023, Migration genetics take flight: genetic and genomic insights into monarch butterfly migration, Current Opinion in Insect Science, 59: 101079.
https://doi.org/10.1016/j.cois.2023.101079

Green D.A., and Kronforst M.R., 2019, Monarch butterflies use an environmentally sensitive, internal timer to control overwintering dynamics, Molecular Ecology, 28(16): 3642-3655.
https://doi.org/10.1111/mec.15178

Guerra P.A., Merlin C., Gegear R.J., and Reppert S.M., 2012, Discordant timing between antennae disrupts sun compass orientation in migratory monarch butterflies, Nature Communications, 3: 958.
https://doi.org/10.1038/ncomms1965

Hammer T.J., McMillan W.O., and Fierer N., 2014, Metamorphosis of a butterfly-associated bacterial community, PLOS ONE, 9(1): e86995.
https://doi.org/10.1371/journal.pone.0086995

Hemstrom W., Freedman M., Zalucki M.P., and Miller M., 2025, Novel genetic association with migratory diapause in Australian monarch butterflies, BMC Ecology and Evolution, 25(1): 43.
https://doi.org/10.1186/s12862-025-02384-w

Hoogshagen M., Hastings A.P., Chavez J., Duckett M., Pettit R., Pahnke A.P., Agrawal A.A., and de Roode J.C., 2024, Mixtures of milkweed cardenolides protect monarch butterflies against parasites, Journal of Chemical Ecology, 50(1): 52-62.
https://doi.org/10.1007/s10886-023-01461-y

Höök L., Vila R., Wiklund C., and Backström N., 2024, Temporal dynamics of faster neo-Z evolution in butterflies, Evolution, 78(9): 1554-1567.
https://doi.org/10.1093/evolut/qpae082

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., and 1000 Genome Project Data Processing Subgroup, 2009, The sequence alignment/map format and SAMtools, Bioinformatics, 25(16): 2078-2079.
https://doi.org/10.1093/bioinformatics/btp352

Livraghi L., Hanly J.J., Loh L.S., Henry A., Keck C., Shirey V.M., Tsai C.C., Yu N.F., Belleghem S.M.V., and Roberts W.M., et al., 2025, Genetic basis of an adaptive polymorphism controlling butterfly silver iridescence, Current Biology, 35(9): 2154-2163.
https://doi.org/10.1016/j.cub.2025.03.028

López-Goldar X., Zhang X., Hastings A.P., Duplais C., and Agrawal A.A., 2024, Plant chemical diversity enhances defense against herbivory, Proceedings of the National Academy of Sciences, 121(51): e2417524121.
https://doi.org/10.1073/pnas.2417524121

Malcolm S.B., and Brower L.P., 1989, Evolutionary and ecological implications of cardenolide sequestration in the monarch butterfly, Experientia, 45(3): 284-295.
https://doi.org/10.1007/BF01951814

Markert M.J., Zhang L., and Reed R.D., 2016, TALEN and CRISPR/Cas9 genome editing in the monarch butterfly, Genome Research, 26(5): 792-801.

Martin A., Papa R., Nadeau N.J., Hill R.I., Counterman B.A., Halder G., Jiggins C.D., Kronforst M.R., Long A.D., and McMillan W.O., et al., 2012, Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand, Proceedings of the National Academy of Sciences, 109(31): 12632-12637.
https://doi.org/10.1073/pnas.1204800109

Merlin C., Gegear R.J., and Reppert S.M., 2009, Antennal circadian clocks coordinate sun compass orientation in migratory monarch butterflies, Science, 325(5948): 1700-1704.
https://doi.org/10.1126/science.1176221

MonarchBase Team, 2012, MonarchBase: the monarch butterfly genome database, Database, 2012: bas028.
https://doi.org/10.1093/database/bas028

Mongue A.J., Hansen M.E.B., Gu L., Sorenson C.E., Walters J.R., and Kronforst M.R., 2017, A neo-sex chromosome in the monarch butterfly, *Danaus plexippus*, G3: Genes, Genomes, Genetics, 7(10): 3281-3294.
https://doi.org/10.1534/g3.117.300187

Mongue A.J., Kuperus P., and Groot A.T., 2025, Genome sequencing confirms cryptic diversity and potentially adaptive changes in gene content of Lepidoptera-infecting Apicomplexa, bioRxiv, 2025-05.
https://doi.org/10.1101/2025.05.29.656805

Mora P., Hospodářská M., Voleníková A.C., Koutecký P., Štundlová J., Dalíková M., Walters J.R., and Nguyen P., 2024, Sex-biased gene content is associated with sex chromosome turnover in Danaini butterflies, Molecular Ecology, 33(24): e17256.
https://doi.org/10.1111/mec.17256

Müller-Theissen M.L., Gottdenker N.L., and Altizer S.M., 2025, Resistance and tolerance to imperfectly specialized parasites: milkweed butterflies and their protozoan parasites, Ecology and Evolution, 15(3): e70979.
https://doi.org/10.1002/ece3.70979

Oberhauser K.S., Prysby M.D., Mattila H.R., Stanley-Horn D.E., Sears M.K., Dively G., Olson E., Pleasants J.M., Lam W.F., and Hellmich R.L., 2001, Temporal and spatial overlap between monarch larvae and corn pollen, Proceedings of the National Academy of Sciences, 98(21): 11913-11918.
https://doi.org/10.1073/pnas.211234298

Petschenka G., and Agrawal A.A., 2015, Milkweed butterfly resistance to plant toxins is linked to sequestration, not coping with toxicity, Proceedings of the Royal Society B: Biological Sciences, 282(1802): 20151865.
https://doi.org/10.1098/rspb.2015.1865

Petschenka G., Fandrich S., Sander N., Wagschal V., Boppré M., and Dobler S., 2013, Stepwise evolution of resistance to toxic cardenolides via genetic substitutions in the Na^+/K^+-ATPase of insects, Proceedings of the National Academy of Sciences, 110(27): 10906-10911.

Rendón-Salinas E., Alonso A., García-Serrano E., Valera-Bermejo A., and Quesada M., 2023, The monarch butterfly in Mexico: a conservation model, Current Opinion in Insect Science, 60: 101112.
https://doi.org/10.1016/j.cois.2023.101112

Sanaei E., Chavez J., Harris E.V., Alcaide T.Y., Baffour-Addo K., Bugay M.J., Adams K.L., Zelaya A., and de Roode J.C., et al., 2024, Microbiome analysis of monarch butterflies reveals effects of development and diet, FEMS Microbiology Ecology, 100(12): fiae143.
https://doi.org/10.1093/femsec/fiae143

Satterfield D.A., Maerz J.C., and Altizer S., 2015, Loss of migratory behaviour increases infection risk for a butterfly host, Proceedings of the Royal Society B: Biological Sciences, 282(1801): 20141734.
https://doi.org/10.1098/rspb.2014.1734

Semmens B.X., Semmens D.J., Thogmartin W.E., Wiederholt R., López-Hoffman L., Diffendorfer J.E., and Pleasants J.M., 2016, Quasi-extinction risk and population targets for the eastern migratory monarch butterfly (*Danaus plexippus*), Scientific Reports, 6: 23265.
https://doi.org/10.1038/srep23265

Shen K., 2024, The use of CRISPR-Cas9 reveals the major genes that control wing patterning and coloration in Nymphalidae butterflies.

Thogmartin W.E., Wiederholt R., Oberhauser K., Drum R.G., Diffendorfer J.E., Altizer S., Taylor O.R., Pleasants J., Semmens D., and Semmens B., et al., 2017, Monarch butterfly population decline in North America: identifying the threatening processes, Royal Society Open Science, 4(9): 170760.
https://doi.org/10.1098/rsos.170760

U.S. Fish and Wildlife Service, 2020, Species status assessment report for the monarch butterfly (*Danaus plexippus*), U.S. Department of the Interior.

Uffelmann E., Huang Q.Q., Munung N.S., De Vries J., Okada Y., Martin A.R., Martin H.C., Lappalainen T., and Posthuma D., 2021, Genome-wide association studies, Nature Reviews Methods Primers, 1(1): 59.
https://doi.org/10.1038/s43586-021-00056-9

Zhan S., Merlin C., Boore J.L., and Reppert S.M., 2011, The monarch butterfly genome yields insights into long-distance migration, Cell, 147(5): 1171-1185.
https://doi.org/10.1016/j.cell.2011.09.052

Zhan S., and Reppert S.M., 2013, MonarchBase: a genomic resource for the monarch butterfly, Nucleic Acids Research, 41(D1): https://doi.org/10.1093/nar/gks1057

Zhan S., Reppert S.M., and Kronforst M.R., 2020, Chromosome-level genome assembly of the monarch butterfly (*Danaus plexippus*), GigaScience, 9(5): giaa042.

Zhan S., Zhang W., Niitepold K., Hsu J., Haeger J.F., Zalucki M.P., Altizer S., de Roode J.C., Reppert S.M., and Kronforst M.R., 2014, The genetics of monarch butterfly migration and warning colouration, Nature, 514(7522): 317-321. https://doi.org/10.1038/nature13812

Zhang L., and Reed R.D., 2016, Genome editing in butterflies reveals that spalt promotes and Distal-less represses eyespot colour patterns, Nature Communications, 7: 11769. https://doi.org/10.1038/ncomms11769

Box 1 Conservation Implications of Monarch Genomics and Metabolomics

Recent advances in monarch genomics and metabolomics provide an opportunity to move beyond abundance-based conservation toward strategies that preserve functional, adaptive capacity across the migratory range.

1. Cheically informed habitat restoration. Milkweed species and populations differ substantially in cardenolide composition, inducibility, and secondary metabolite diversity, with direct consequences for monarch survival, chemical defense, and resistance to the protozoan parasite *Ophryocystis elektroscirrha* (OE) (de Roode et al., 2008; Agrawal et al., 2012; Petschenka and Agrawal, 2015). Metabolomic screening of milkweeds used in restoration projects could identify plant chemotypes that optimize larval performance while enhancing parasite resistance and predator deterrence, avoiding one-size-fits-all planting strategies.

2. Preserving adaptive genetic variation for migration. Population genomic studies indicate that migration, diapause, lipid storage, and navigation are polygenic traits shaped by subtle allele-frequency shifts across many loci (Zhan et al., 2014; Freedman and Kronforst, 2023). Conservation actions that maintain connectivity among breeding, migratory, and overwintering regions are therefore essential to preserve adaptive alleles associated with circadian timing, endocrine regulation, and metabolic endurance.

3. Metaolites as early-warning indicators. Metabolomic profiles-such as cardenolide sequestration patterns, lipid reserves, and stress-response metabolites-may provide sensitive indicators of physiological condition and migratory readiness before population declines are detectable through census data alone (Semmens et al., 2016; Thogmartin et al., 2017). Integrating metabolomic monitoring into long-term surveys could improve detection of sublethal stress caused by climate extremes, pesticide exposure, or host-plant mismatch.

4. Integrating host–parasite–microbiome dynamics. Chemical defense, parasite resistance, and gut microbiome composition interact to shape monarch fitness. Restoration strategies that consider host-plant chemistry alongside disease pressure and microbial interactions may reduce parasite prevalence and improve survival during migration and overwintering (de Roode et al., 2008; Hammer et al., 2014).

5. From molecular insight to management practice. By integrating causal genomics, pan-genomics, regulatory neurogenomics, and metabolomics, conservation efforts can prioritize not only habitat quantity but also genetic, chemical, and physiological quality. Such mechanistically informed strategies are likely to be more robust to environmental change, supporting long-term persistence of migratory monarch populations in an increasingly variable landscape (U.S. Fish and Wildlife Service, 2020; Erickson et al., 2023).

International Journal of Molecular Ecology and Conservation, 2026, Vol.16, No.1, 13-21
http://ecoevopublisher.com/index.php/ijmec

Research Insight Open Access

Integrating Ecology and Genomics to Understand Population Dynamics and Adaptive Evolution in the Saker Falcon

Xuming Lyu, Yeping Han
Institute of Life Sciences, Jiyang Colloge of Zhejiang AandF University, Zhuji, 311800, Zhejiang, China
Corresponding author: yeping.han@jicat.org

International Journal of Molecular Ecology and Conservation, 2026, Vol.16, No.1 doi: 10.5376/ijmec.2026.16.0002
Received: 09 Jan., 2026
Accepted: 07 Feb., 2026
Published: 27 Feb., 2026

Preferred citation for this article:
Lyu X.M., and Han Y.P., 2026, Integrating ecology and genomics to understand population dynamics and adaptive evolution in the Saker Falcon, International Journal of Molecular Ecology and Conservation, 16(1): 13-21 (doi: 10.5376/ijmec.2026.16.0002)

Abstract This study systematically integrates the latest findings from field-based ecological surveys and genomic research to analyze the population dynamics and adaptive evolution of the Saker Falcon at both global and regional scales. The research encompasses ecological characteristics such as geographic distribution, breeding biology, diet composition, and migration patterns, as well as genomic findings on genetic structure, gene flow, signals of selection, and local adaptation. A case study comparing Mongolian and Central European populations reveals adaptive divergence shaped by distinct ecological environments and discusses its implications for conservation strategy development. The study aims to guide future Saker Falcon conservation research in optimizing monitoring frameworks, applying genomic technologies, and advancing adaptive management.

Keywords Saker Falcon; Population dynamics; Adaptive evolution; Genomics; Conservation biology

1 Introduction

The Saker Falcon (*Falco cherrug*) is a large, wide-ranging raptor distributed across Eurasia, from Central Europe to East Asia, and is recognized for its ecological role as a top predator in grassland and steppe ecosystems (Stretesky et al., 2018). Despite its broad range, the species has experienced dramatic population declines over recent decades, primarily due to habitat loss, prey depletion, electrocution, and unsustainable trapping for the global falconry trade, leading to its current classification as Endangered by the IUCN (Stretesky et al., 2018; Karyakin et al., 2023; Kovacs et al., 2023). Conservation efforts have been implemented in several countries, including artificial nest provision, habitat management, and legal protections, with varying degrees of success (Bagyura et al., 2023; Hohenegger, 2023; Zhang et al., 2024). Understanding the population dynamics and adaptive evolution of the Saker Falcon is crucial for effective conservation, as these processes underpin the species' resilience to environmental change and anthropogenic pressures (Zhan et al., 2013; Zinevich et al., 2023).

Integrating ecological field data with genomic analyses offers a comprehensive approach to raptor conservation. Field observations provide essential information on population size, breeding success, migration routes, and threats such as electrocution and poaching (Dixon et al., 2020; Karyakin et al., 2023; Zhang et al., 2024). Genomic data, including whole-genome sequencing and population genetic studies, reveal patterns of genetic diversity, population structure, hybridization, and adaptive traits that are not detectable through fieldwork alone (Nittinger et al., 2007; Zhan et al., 2013; Zinevich et al., 2023; Petrov et al., 2024). For the Saker Falcon, genomic studies have clarified taxonomic uncertainties, identified distinct lineages and ecotypes, and uncovered signatures of rapid evolution related to predatory adaptations and environmental stressors (Zhan et al., 2013; Zinevich et al., 2023; Petrov et al., 2024; Al-Ajli et al., 2025). The complementary use of these data sources enhances our understanding of migration, gene flow, and local adaptation, informing targeted management and reintroduction strategies (Zhan et al., 2013; Petrov et al., 2023; Zinevich et al., 2023; Petrov et al., 2024; Zhang et al., 2024).

This study aims to integrate the latest findings from field-based ecological research and genomic studies to assess global and regional population trends and the drivers of decline or recovery, analyze the genetic structure, diversity, and adaptive traits within and among populations, and propose an adaptive management framework to provide evidence-based support for the long-term conservation of the Saker Falcon.

2 Ecology and Life History of the Saker Falcon

2.1 Geographic distribution and habitat preferences

The Saker Falcon (Falco cherrug) has a broad Palearctic distribution, spanning from Central and Eastern Europe through Central Asia to western China and Mongolia (Karyakin et al., 2022; Karyakin et al., 2023). Populations are found in diverse habitats, including arid steppes, grasslands, semi-deserts, and mountainous regions. In areas such as the Karatau Mountains (Kazakhstan), suitable breeding habitats have been mapped to over 4,200 km², with total habitats exceeding 9,000 km² (Karyakin et al., 2022). In the Sanjiangyuan National Park (China), habitat suitability is strongly influenced by elevation and temperature, with Saker Falcons favoring open landscapes and showing sensitivity to extreme temperatures (Zhang et al., 2019). Human disturbance is less significant in remote regions, but habitat changes due to land use and prey availability can drive shifts in nesting sites, as observed in Slovakia where populations moved from highlands to lowlands in response to forest management and prey decline (Chavko et al., 2019).

2.2 Breeding biology and reproductive strategies

Saker Falcons typically breed from April to July, with clutch sizes ranging from three to five eggs (average 4.0) (Yi-Qun et al., 2007). They often utilize nests built by other large birds, such as buzzards, eagles, or ravens, and increasingly nest on artificial structures like power poles and nest boxes (Chavko et al., 2019; Zhatkanbaev et al., 2023). Breeding success is closely linked to prey abundance and climatic conditions. In Mongolia, artificial nest provision has supported large managed populations, with breeding density and fledging success positively correlated with small mammal prey and favorable weather (Zhang et al., 2024). In Xinjiang, China, food availability is a key determinant of clutch size and fecundity, and nest success rates can exceed 80% under optimal conditions (Yi-Qun et al., 2007). Instances of cainism (sibling aggression) and nest abandonment have been documented, often associated with food scarcity or disturbance (Zhatkanbaev et al., 2023; Bold et al., 2024).

2.3 Diet composition and hunting behavior

The Saker Falcon is an opportunistic predator, with diet composition varying by region and prey availability. In Central Asia and Kazakhstan, small mammals such as the Great Gerbil (*Rhombomys opimus*) and other rodents are primary prey, but birds (e.g., pigeons, starlings, crows) become more important when rodent populations decline (Chavko et al., 2019; Karyakin et al., 2022; Zhatkanbaev et al., 2023). In Slovakia, long-term studies show a shift from mammals to birds in the diet, with domestic pigeons comprising up to 62% of prey in some areas, and mammals like voles and ground squirrels declining due to habitat changes (Chavko et al., 2014; Chavko et al., 2019). Saker Falcons also hunt reptiles, fish, and occasionally feed on carrion, demonstrating flexible foraging strategies (Zhatkanbaev et al., 2023). Prey availability directly influences home range size and breeding success, particularly for males during the breeding season (Bold et al., 2024; Zhang et al., 2024).

2.4 Migration patterns and seasonal movements

Saker Falcons exhibit a range of movement strategies, from resident to migratory, depending on geographic location and resource availability. Satellite tracking in Mongolia reveals strong territoriality during breeding, with minimal overlap between neighboring pairs (Bold et al., 2024). Males adjust their home range size in response to prey density, occupying smaller territories in areas with abundant rodents (Bold et al., 2024). Seasonal movements are influenced by prey fluctuations, with some populations displaying nomadism or long-distance migration to wintering grounds in southern Asia or the Middle East (Zhang et al., 2019; Karyakin et al., 2022). In China's Sanjiangyuan National Park, Saker Falcons' wintering home ranges are shaped by environmental variables such as elevation and temperature, and overlap with other raptor species is limited by dietary and spatial preferences (Zhang et al., 2019).

3 Population Dynamics Analysis Based on Field Observations

3.1 Long-term monitoring methodologies

Long-term field studies are essential for understanding population dynamics, as they capture the heterogeneity and temporal variability that drive population processes (Reinke et al., 2019). Common methodologies include satellite tracking, which provides detailed data on individual movements and spatial use; banding (ringing), which

enables the estimation of survival and dispersal rates through mark-recapture analysis; and on-site nest monitoring, which yields direct measures of breeding success and productivity. Integrating these approaches allows for robust, multi-scale insights into population trends and demographic parameters (Zipkin et al., 2017; Reinke et al., 2019). Recent advances also emphasize the value of combining different data types—such as count data and detection-nondetection records—within unified analytical frameworks to improve inference about abundance and demographic rates, even when detection probabilities vary (Buckland et al., 2004; Hostetler and Chandler, 2015; Zipkin et al., 2017).

3.2 Demographic parameters

Key demographic parameters assessed through field observations include survival rates, breeding success rates, and juvenile survival. These metrics are critical for modeling population growth or decline and for identifying life stages most sensitive to environmental pressures. State-space models and integrated population models are increasingly used to estimate these parameters, accounting for both ecological process variation and observation error (Buckland et al., 2004; Hostetler and Chandler, 2015; Zipkin et al., 2017). Such models can incorporate data from marked and unmarked individuals, providing more accurate estimates of survival and reproduction over time (Buckland et al., 2004; Zipkin et al., 2017).

3.3 Population trends and threats

Long-term monitoring reveals that population trends are shaped by a combination of natural and anthropogenic factors. Habitat loss and fragmentation remain primary threats, reducing available nesting and foraging sites. Illegal trade, particularly in high-value raptor species, can cause significant population declines. Environmental pollution, including pesticides and heavy metals, further impacts survival and reproductive success. Field-based population viability analyses, especially when integrated with remote sensing and landscape data, help forecast the effects of these threats and guide conservation actions (Reinke et al., 2019; Giezendanner et al., 2020).

3.4 Influence of climatic and anthropogenic factors

Climatic variables such as temperature and precipitation directly influence demographic rates by affecting food availability, breeding timing, and survival (Giezendanner et al., 2020; Neta et al., 2021). Anthropogenic factors—including land use change, urbanization, and direct persecution—can exacerbate natural fluctuations, leading to increased extinction risk. Advanced modeling frameworks now incorporate both static (e.g., topography) and dynamic (e.g., climate, vegetation) variables to predict spatial and temporal trends in population occupancy and viability (Giezendanner et al., 2020; Neta et al., 2021). These approaches enable near-term ecological forecasting and support adaptive management in the face of rapid environmental change (Reinke et al., 2019; Giezendanner et al., 2020; Neta et al., 2021).

4 Genomic Approaches to Studying Adaptive Evolution

4.1 Genomic sequencing and assembly strategies

Advances in high-throughput sequencing technologies have enabled the generation of whole-genome assemblies for both model and non-model organisms, providing the foundation for comparative and population genomics studies of adaptive evolution (Bomblies and Peichel, 2022; Hu et al., 2023). These strategies include the use of next-generation sequencing to capture genome-wide variation, allowing for the identification of both single nucleotide polymorphisms (SNPs) and structural variants such as gene duplications, deletions, and transposable element insertions (Villanueva-Cañas et al., 2017; Bomblies and Peichel, 2022). The increasing accessibility of genomic data facilitates the detection of gene loss events and the construction of high-quality gene catalogs, which are crucial for understanding the molecular basis of adaptation (Villanueva-Cañas et al., 2017; Sharma et al., 2018).

4.2 Population genetic structure and gene flow

Population genomics enables the analysis of genetic structure and gene flow within and between populations, which is essential for understanding the evolutionary processes shaping adaptive traits (González-Martínez et al., 2006; Combrink et al., 2024). By sampling many individuals across the species' range, researchers can infer patterns of hybridization, introgression, and the re-use of standing genetic variation during adaptation (Combrink

et al., 2024). These analyses reveal how gene flow and population connectivity contribute to the maintenance or erosion of adaptive genetic diversity, informing conservation strategies for species with fragmented or declining populations (Harrisson et al., 2014; Combrink et al., 2024).

4.3 Signals of selection and adaptive traits

Detecting signals of selection involves identifying genomic regions or loci that show evidence of positive selection, often through statistical methods such as the McDonald-Kreitman test, codon substitution models, or environmental association analyses (Williamson et al., 2007; Villanueva-Cañas et al., 2017; Huang, 2020). Genes associated with high-altitude adaptation, migratory ability, and hunting efficiency can be pinpointed by linking genotype to phenotype using comparative genomics and functional assays (Orteu and Jiggins, 2020; Bomblies and Peichel, 2022; Hu et al., 2023). For example, highly expressed genes and metabolic genes have been shown to exhibit higher rates of adaptation, and structural variants like transposable elements may also contribute to adaptive evolution (Villanueva-Cañas et al., 2017; Huang, 2020). The integration of these approaches allows for the identification of both large-effect mutations and polygenic adaptation underlying complex traits (Huang, 2020; Bomblies and Peichel, 2022; Hu et al., 2023).

4.4 Integration of genomic data with ecological insights

A holistic understanding of adaptation requires the integration of genomic data with ecological and field-based observations (Harrisson et al., 2014; Bomblies and Peichel, 2022). This interdisciplinary approach connects genetic variants to phenotypic traits and fitness in natural environments, enabling the study of adaptation in the context of real-world ecological pressures (Harrisson et al., 2014; Bomblies and Peichel, 2022; Hu et al., 2023). Genomic estimates of evolutionary potential, when combined with ecological data, provide robust predictions of population persistence and inform adaptive management strategies in conservation biology (Harrisson et al., 2014; Hu et al., 2023).

5 Case Study: Adaptive Divergence in Mongolian and Central European Populations

5.1 Case background

Mongolian and Central European Saker Falcon populations inhabit ecologically distinct regions. Mongolian populations are found in expansive grasslands and steppe environments characterized by pastoralism, open landscapes, and variable climates, while Central European populations occupy farmlands and fragmented habitats shaped by intensive agriculture and human settlement (Jeong et al., 2020; Yang et al., 2021). These ecological differences influence resource availability, predator-prey dynamics, and exposure to environmental pressures, setting the stage for divergent adaptive strategies.

5.2 Field observation data comparison

Field studies reveal notable differences in habitat density, prey composition, and breeding success between the two regions. Mongolian Saker Falcons benefit from vast, contiguous habitats with high densities of small mammal prey, supporting larger home ranges and stable breeding populations. In contrast, Central European populations contend with fragmented landscapes, lower prey diversity, and greater anthropogenic disturbance, often resulting in reduced breeding success and altered foraging behavior. These ecological contrasts are reflected in population density, reproductive output, and survival rates.

5.3 Genomic evidence of local adaptation

Genomic analyses highlight significant genetic divergence and local adaptation between Mongolian and Central European populations. Mongolian populations exhibit high genetic diversity and distinct genetic clusters, shaped by historical admixture with both Eastern and Western Eurasian ancestries (Figure 1) (Jeong et al., 2020; Derenko et al., 2021; Yang et al., 2021). Signals of selection have been detected in genes related to metabolic rate, immune function (notably the MHC region), and environmental tolerance, reflecting adaptation to the harsh, variable climates of the steppe (Yang et al., 2021). While specific studies on Saker Falcons are limited, research on Mongolian populations more broadly suggests that adaptive traits such as plumage coloration, metabolic efficiency, and climate resilience are under selection, supporting local adaptation to regional ecological conditions (Jeong et al., 2020; Yang et al., 2021).

5.4 Implications for conservation strategies

The observed adaptive divergence underscores the need for region-specific conservation strategies. For Mongolian populations, maintaining large, connected habitats and supporting traditional pastoral land use are critical for preserving genetic diversity and adaptive potential. In Central Europe, conservation should focus on mitigating habitat fragmentation, enhancing prey availability, and reducing anthropogenic pressures. Genetic monitoring and the integration of genomic data into management plans will help safeguard locally adapted lineages and ensure the long-term viability of Saker Falcon populations across their range (Jeong et al., 2020; Derenko et al., 2021; Yang et al., 2021).

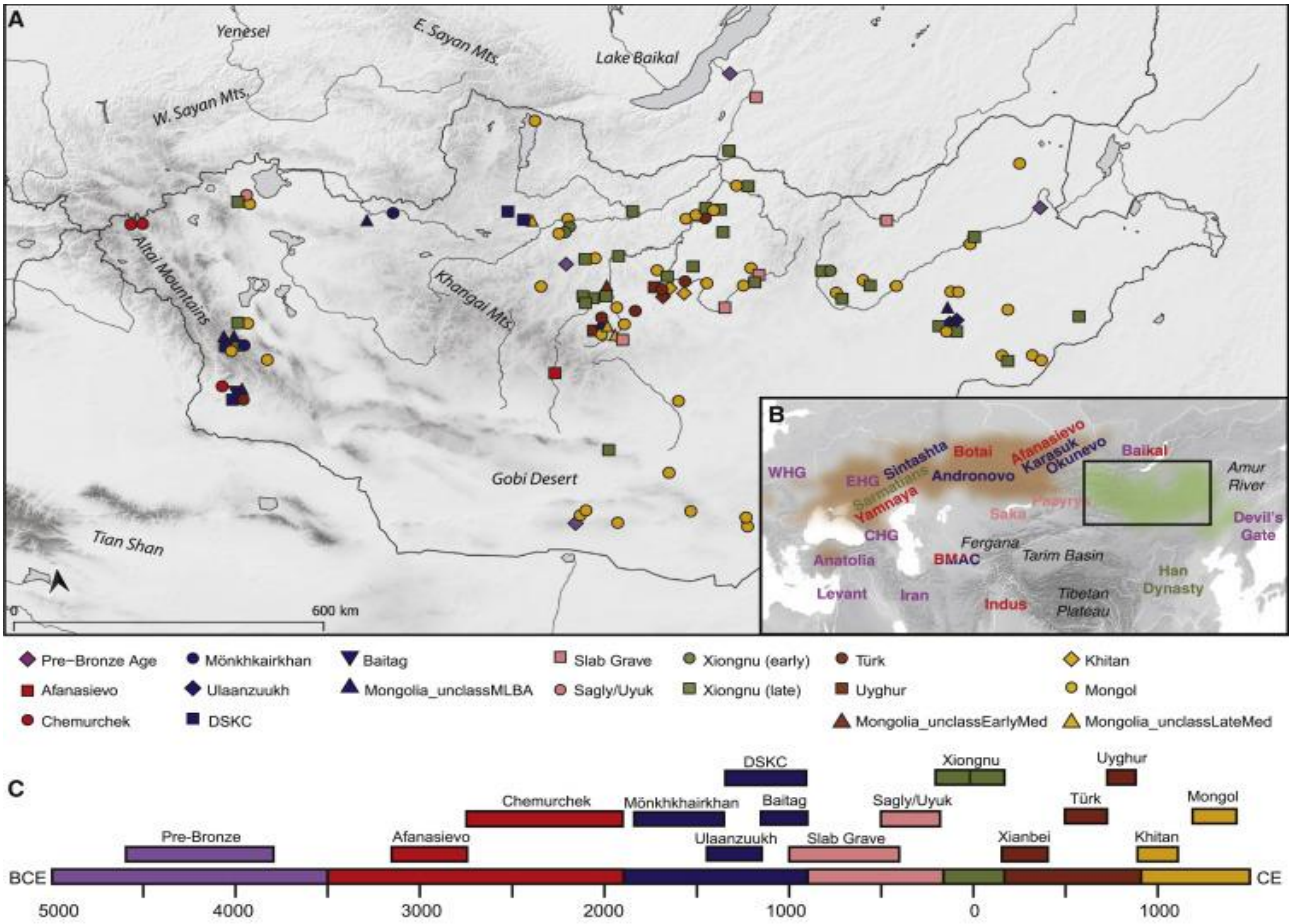

Figure 1 Overview of ancient populations and time periods (adopted from Jeong et al., 2020)

Image caption: (A) Distribution of sites with their associated culture and time period indicated by color: Pre-Bronze, purple; Early Bronze, red; Middle/Late Bronze, blue; Early Iron, pink; Xiongnu, green; Early Medieval, brown; Late Medieval, gold (see STAR Methods). See Figure S1A and Table S1B for site codes and labels; (B) Inset map of Eurasia indicating area of present study (box) and the locations of other ancient populations referenced in the text, colored by time period. The geographic extent of the Western/Central Steppe is indicated in light brown, and the Eastern Steppe is indicated in light green; (C) Timeline of major temporal periods and archaeological cultures in Mongolia. Site locations have been jittered to improve visibility of overlapping sites (Adopted from Jeong et al., 2020)

6 Integrative Analysis: Linking Ecology and Genomics

6.1 Correlating environmental variables with genetic markers

Ecological genomics leverages functional genomic approaches to identify genes and genomic regions associated with responses to specific environmental variables, such as temperature, habitat type, or resource availability (Ungerer et al., 2008; Katsikis et al., 2014). By analyzing correlations between environmental gradients and genetic markers, researchers can uncover the genetic basis of ecologically relevant phenotypic variation and adaptive traits (Ungerer et al., 2008; Katsikis et al., 2014). Statistical methods, such as multivariate analyses and environmental association studies, are commonly used to link environmental metadata with genomic data,

allowing for the detection of loci under selection in response to ecological pressures (Pérez-Cobas et al., 2020; Ozerov et al., 2025). This integrative approach enhances the understanding of how natural selection operates in heterogeneous environments and informs the identification of adaptive genetic variation.

6.2 Identifying eco-genomic units for management

Defining eco-genomic units—populations or lineages characterized by distinct ecological and genomic profiles—enables more precise conservation management (Katsikis et al., 2014; Guevara-Escudero et al., 2021). Integrative studies that combine ecological, phylogeographic, and genomic data can map the geographic distribution of genealogical lineages and adaptive traits across landscapes (Guevara-Escudero et al., 2021). This process helps identify locally adapted populations and informs the delineation of management units that reflect both genetic diversity and ecological function, which is critical for maintaining evolutionary potential and ecosystem resilience (Katsikis et al., 2014; Guevara-Escudero et al., 2021).

6.3 Predictive models for population viability

The integration of ecological and genomic data supports the development of predictive models for population viability under changing environmental conditions (Richardson et al., 2016; Matthews et al., 2018). By incorporating genomic estimates of adaptive capacity and ecological variables, these models can forecast population responses to threats such as habitat loss, climate change, and disease (Richardson et al., 2016; Matthews et al., 2018). Such models are essential for adaptive management, as they allow conservationists to anticipate future challenges and prioritize actions that enhance the persistence of genetically and ecologically important populations (Richardson et al., 2016; Matthews et al., 2018).

7 Conservation and Management Implications

7.1 Conservation priorities based on integrated data

Effective conservation of the Saker Falcon requires integrating ecological, genomic, and social data to set priorities that address both species persistence and ecosystem health. Area-based conservation remains foundational, but its effectiveness depends on adaptive management, robust monitoring, and the use of open data infrastructures to track population trends and threats (Maxwell et al., 2020; Hoffmann, 2021). Conservation actions—such as habitat protection, invasive species control, and restoration—have been shown to improve or slow declines in biodiversity in most cases, but require scaling up and continuous evaluation to meet global targets (Maxwell et al., 2020; Langhammer et al., 2024). Prioritizing conservation actions should also consider the evolutionary impacts of management, ensuring that strategies maintain genetic diversity and adaptive potential (Shefferson et al., 2018).

7.2 Transboundary cooperation in falcon protection

Given the Saker Falcon's wide range across multiple countries, transboundary cooperation is essential for effective conservation. International agreements, shared monitoring protocols, and coordinated management of protected areas can help address threats such as habitat loss, illegal trade, and environmental change that cross national borders (Van Kerkhoff et al., 2018; Maxwell et al., 2020). Collaborative frameworks should secure adequate financing, harmonize biodiversity policies, and mainstream conservation into broader land, water, and sea management to ensure long-term success (Van Kerkhoff et al., 2018; Maxwell et al., 2020). The "One Conservation" approach, which integrates in situ and ex situ efforts and involves multiple sectors, further highlights the need for joint action across regions and disciplines (Pizzutto et al., 2021).

7.3 Role of citizen science and local communities

Engaging local communities and citizen scientists is critical for the long-term success of conservation initiatives. Community-based conservation, co-management, and biocultural approaches that integrate local knowledge and address social, economic, and cultural needs can reduce conflicts and increase support for protected areas (Bennett, 2016; He et al., 2020; Hoffmann, 2021). Positive perceptions and active participation by local people enhance compliance, monitoring, and adaptive management, while also ensuring that conservation benefits are equitably shared (Bennett, 2016; He et al., 2020; Mubalama et al., 2020). Citizen science initiatives can fill data gaps, improve monitoring efficiency, and foster stewardship, making them valuable tools for both research and management (Bennett, 2016; Hoffmann, 2021).

8 Challenges and Future Directions in Saker Falcon Research

8.1 Data integration limitations and biases

Integrating field observations with genomic data presents significant challenges, including inconsistencies in data collection methods, spatial and temporal mismatches, and varying data quality. These limitations can introduce biases that affect the reliability of ecological-genomic analyses. For example, differences in monitoring intensity or technology adoption across regions may lead to uneven data coverage, while the integration of heterogeneous datasets requires robust frameworks to ensure comparability and minimize error propagation. Addressing these challenges will require standardized protocols, improved data sharing infrastructures, and interdisciplinary collaboration to harmonize methodologies and reduce integration biases.

8.2 Emerging genomic technologies in wildlife conservation

Rapid advances in genomic technologies—such as next-generation sequencing, environmental DNA (eDNA) analysis, and portable sequencing platforms—are transforming wildlife conservation. These tools enable high-resolution population genetic studies, real-time monitoring of genetic diversity, and the detection of adaptive genetic variation even in non-model species. However, the adoption of these technologies also brings challenges, including the need for specialized expertise, high costs, and the management of large, complex datasets. Future directions should focus on making genomic tools more accessible, developing user-friendly analytical pipelines, and integrating genomic insights into practical conservation management.

8.3 Long-term monitoring and climate change adaptation

Long-term ecological monitoring remains essential for understanding population trends and adaptive responses to climate change. However, sustaining such efforts is challenged by funding limitations, logistical constraints, and the need for consistent methodologies over time. Climate change introduces additional complexity, as shifting environmental baselines may alter species distributions, phenology, and adaptive pressures. Future research should prioritize the development of adaptive monitoring frameworks that can respond to changing conditions, leverage remote sensing and automated data collection, and incorporate predictive modeling to inform proactive conservation strategies.

Acknowledgments

The EcoEvo Publiser extends sincere thanks to two anonymous peer reviewers for their feedback on the manuscript.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abdellaoui A., Yengo L., Verweij K., and Visscher P., 2023, 15 years of GWAS discovery: Realizing the promise, American Journal of Human Genetics, 110(2): 179-194.
https://doi.org/10.1016/j.ajhg.2022.12.011

Alqudah A.M., Sallam A., Baenziger P.S., and Börner A., 2020, GWAS: fast-forwarding gene identification and characterization in temperate cereals: lessons from barley-a review, J. Adv. Res., 22: 119-135.
https://doi.org/10.1016/j.jare.2019.10.013

Ashwath M., Lavale S., Santhoshkumar A., Mohapatra S., Bhardwaj A., Dash U., Shiran K., Samantara K., and Wani S., 2023, Genome-wide association studies: an intuitive solution for SNP identification and gene mapping in trees, Funct. Integr. Genomics, 23(4): 297.
https://doi.org/10.1007/s10142-023-01224-8

Bashir L., Mehmood A., Manzoor S., Thendral U., Yadav J., Saha S., Yadav M., Meena D., and Padder U., 2024, A comprehensive review on GWAS: basic concepts and role in agriculture, Int. J. Agric. Ext. Soc. Dev., 7(10): 302-314.
https://doi.org/10.33545/26180723.2024.v7.i10e.1231

Baxter I., 2020, We aren't good at picking candidate genes, and it's slowing us down, Curr. Opin. Plant Biol., 54: 57-60.
https://doi.org/10.1016/j.pbi.2020.01.006

Cano-Gamez E., and Trynka G., 2020, From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases, Front. Genet., 11: 424.
https://doi.org/10.3389/fgene.2020.00424

Chang-Brahim I., Koppensteiner L., Beltrame L., Bodner G., Saranti A., Salzinger J., Fanta-Jende P., Sulzbachner C., Bruckmüller F., Trognitz F., Samad-Zamini M., Zechner E., Holzinger A., and Molin E., 2024, Reviewing the essential roles of remote phenotyping, GWAS and explainable AI in practical marker-assisted selection for drought-tolerant winter wheat breeding, Front. Plant Sci., 15: 1319938.
https://doi.org/10.3389/fpls.2024.1319938

Chen G.B., 2014, Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression, Front. Genet., 5: 107.
https://doi.org/10.3389/fgene.2014.00107

Chen G.B., 2016, On the reconciliation of missing heritability for genome-wide association studies, Eur. J. Hum. Genet., 24(12): 1810-1816.
https://doi.org/10.1038/ejhg.2016.89

Fang X.J., Wu W.R., and Tang J.L. (eds.), 2001, Crop DNA marker-assisted breeding, Science Press, Beijing, China, pp. 1-84.

Feingold E., 2002, Regression-based quantitative-trait-locus mapping in the 21st century, Am. J. Hum. Genet., 71(2): 217-222.
https://doi.org/10.1086/341964

Gallagher M.D., and Chen-Plotkin A.S., 2018, The post-GWAS era: From association to function, Am. J. Hum. Genet., 102(5): 717-730.
https://doi.org/10.1016/j.ajhg.2018.04.002

He J., Meng S., Zhao T., Xing G., Yang S., Li Y., Guan R., Lu J., Wang Y., Xia Q., Yang B., and Gai J., 2017, An innovative procedure of genome-wide association analysis fits studies on germplasm population and plant breeding, Theor. Appl. Genet., 130(12): 2327-2343.
https://doi.org/10.1007/s00122-017-2962-9

Ibrahim A., Zhang L., Niyitanga S., Afzal M., Xu Y., Zhang L., Zhang L., and Qi J., 2020, Principles and approaches of association mapping in plant breeding, Trop. Plant Biol., 13(1): 212-224.
https://doi.org/10.1007/s12042-020-09261-4

Jeong M., Pazokitoroudi A., Liu Z., and Sankararaman S., 2024, Scalable summary-statistics-based heritability estimation method with individual genotype level accuracy, Genome Res., 34(7): 1286-1293.
https://doi.org/10.1101/gr.279207.124

Kushanov F., Turaev O., Ernazarova D., Gapparov B., Oripova B., Kudratova M., Rafieva F., Khalikov K., Erjigitov D., Khidirov M., Kholova M., Khusenov N., Amanboyeva R., Saha S., Yu J., and Abdurakhmonov I., 2021, Genetic diversity, QTL mapping, and marker-assisted selection technology in cotton (Gossypium spp.), Front. Plant Sci., 12: 779386.
https://doi.org/10.3389/fpls.2021.779386

Li C., Li Y., Bradbury P., Wu X., Shi Y., Song Y., Zhang D., Rodgers-Melnick E., Buckler E., Zhang Z., Li Y., and Wang T., 2015, Construction of high-quality recombination maps with low-coverage genomic sequencing for joint linkage analysis in maize, BMC Biol., 13: 78.
https://doi.org/10.1186/s12915-015-0187-4

Liu H., and Chen G., 2022, A novel genomic prediction method combining randomized Haseman-Elston regression with a modified algorithm for Proven and Young for large genomic data, Crop J., 10(2): 550-554.
https://doi.org/10.1016/j.cj.2021.09.001

Marigorta U.M., Rodríguez J.A., Gibson G., and Navarro A., 2018, Replicability and prediction: Lessons and challenges from GWAS, Trends Genet., 34(7): 504-517.
https://doi.org/10.1016/j.tig.2018.03.005

Mei B., and Wang Z., 2016, An efficient method to handle the 'large p, small n'problem for genomewide association studies using Haseman-Elston regression, J. Genet., 95 (4): 847-852.
https://doi.org/10.1007/s12041-016-0705-3

Meng L., Li H., Zhang L., and Wang J., 2015, QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations, Crop J., 3(3): 269-283.
https://doi.org/10.1016/j.cj.2015.01.001

Nandi S., Varotariya K., Luhana S., Kyada A., Saha A., Roy N., Sharma N., and Rambabu D., 2024, GWAS for identification of genomic regions and candidate genes in vegetable crops, Funct. Integr. Genomics, 24(6): 203.
https://doi.org/10.1007/s10142-024-01477-x

Pasaniuc B., and Price A.L., 2016, Dissecting the genetics of complex traits using summary association statistics, Nat. Rev. Genet., 18(2): 117-127.
https://doi.org/10.1038/nrg.2016.142

Pflieger S., Lefebvre V., and Causse M., 2001, The candidate gene approach in plant genetics: A review, Mol. Breed., 7(4): 275-291.
https://doi.org/10.1023/A:1011605013259

Qu P., Shi J., Chen T., Chen K., Shen C., Wang J., Zhao X., Ye G., Xu J., and Zhang L., 2020, Construction and integration of genetic linkage maps from three multi-parent advanced generation inter-cross populations in rice, Rice, 13(1): 13.
https://doi.org/10.1186/s12284-020-0373-z

Raj S., and Nadarajah K., 2022, QTL and candidate genes: Techniques and advancement in abiotic stress resistance breeding of major cereals, Int. J. Mol. Sci., 24(1): 6.
https://doi.org/10.3390/ijms24010006

Sebastiani P., Timofeev N., Dworkis D., Perls T.T., and Steinberg M.H., 2009, Genome‐wide association studies and the genetic dissection of complex traits, Am. J. Hematol., 84(8): 504-515.
https://doi.org/10.1002/ajh.21440

Sham P.C., and Purcell S., 2001, Equivalence between Haseman-Elston and variance-components linkage analyses for sib pairs, Am. J. Hum. Genet., 68(6): 1527-1532.
https://doi.org/10.1086/320593

Sofer T., 2017, Confidence intervals for heritability via Haseman-Elston regression, Stat. Appl. Genet. Mol. Biol., 16(4): 259-273.
https://doi.org/10.1515/sagmb-2016-0076

Stanton-Geddes J., Paape T., Epstein B., Briskine R., Yoder J., Mudge J., Bharti A., Farmer A., Zhou P., Denny R., May G., Erlandson S., Yakub M., Sugawara M., Sadowsky M., Young N., and Tiffin P., 2013, Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in Medicago truncatula, PLoS ONE, 8 (6): e65688.
https://doi.org/10.1371/journal.pone.0065688

Susmitha P., Kumar P., Yadav P., Sahoo S., Kaur G., Pandey M., Singh V., Tseng T., and Gangurde S., 2023, Genome-wide association study as a powerful tool for dissecting competitive traits in legumes, Front. Plant Sci., 14: 1123631.
https://doi.org/10.3389/fpls.2023.1123631

Tam V., Patel N., Turcotte M., Bossé Y., Paré G., and Meyre D., 2019, Benefits and limitations of genome-wide association studies, Nat. Rev. Genet., 20(8): 467-484.
https://doi.org/10.1038/s41576-019-0127-1

Taniguti C., Taniguti L., Amadeu R., Lau J., De Siqueira Gesteira G., De Paula Oliveira T., et al., 2022, Developing best practices for genotyping-by-sequencing analysis in the construction of linkage maps, GigaScience, 12(1): giad092.
https://doi.org/10.1093/gigascience/giad092

Tibbs Cortes L., Zhang Z., and Yu J., 2021, Status and prospects of genome-wide association studies in plants, The plant genome, 14(1): e20077.
https://doi.org/10.1002/tpg2.20077

Uffelmann E., Huang Q.Q., Munung N.S., De Vries J., Okada Y., Martin A.R., Martin H.C., Lappalainen T., and Posthuma D., 2021, Genome-wide association studies, Nat. Rev. Methods Primers, 1(1): 1-21.
https://doi.org/10.1038/s43586-021-00056-9

Wang T., and Elston R.C., 2005, Two-level Haseman-Elston regression for general pedigree data analysis, Genet. Epidemiol., 29(1): 12-22.
https://doi.org/10.1002/gepi.20075

Wang X., Wang J., Xia X., Xu X., Li L., Cao S., Hao Y., and Zhang L., 2024, Effect of genotyping errors on linkage map construction based on repeated chip analysis of two recombinant inbred line populations in wheat (Triticum aestivum L.), BMC Plant Biol., 24(1): 306.
https://doi.org/10.1186/s12870-024-05005-8

Watanabe K., Stringer S., Frei O., Mirkov U.G., de Leeuw C.A., Polderman T.J., et al., 2019, A global overview of pleiotropy and genetic architecture in complex traits, Nat. Genet., 51(9): 1339-1348.
https://doi.org/10.1038/s41588-019-0481-0

Xu T., Qi G., Zhu J., Xu H., and Chen G., 2021, Subsampling technique to estimate variance component for UK-Biobank traits, Front. Genet., 12: 612045.
https://doi.org/10.3389/fgene.2021.612045

Xu Y., Li P., Yang Z., and Xu C., 2017, Genetic mapping of quantitative trait loci in crops, Crop J., 5(2): 175-184.
https://doi.org/10.1016/j.cj.2016.06.003

Zhang L., Li H., and Wang J., 2015, Linkage analysis and map construction in genetic populations of clonal F1 and double cross, G3 (Bethesda), 5(3): 427-439.
https://doi.org/10.1534/g3.114.016022

Zhang Y., Jia Z., and Dunwell J.M., 2019, The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits, Front. Plant Sci., 10: 100.
https://doi.org/10.3389/fpls.2019.00100

Zhang Y., Wang M., Li Z., Yang X., Li K., Xie A., Dong F., Wang S.H., Yan J.B., and Liu J., 2024, An overview of detecting gene-trait associations by integrating GWAS summary statistics and eQTLs, Sci. China Life Sci., 67 (6): 1133-1154.
https://doi.org/10.1007/s11427-023-2522-8

Zheng C., Boer M., and van Eeuwijk F., 2019, Construction of genetic linkage maps in multiparental populations, Genetics, 212(4): 1031-1044.
https://doi.org/10.1534/genetics.119.302229

Zhou X., 2017, A unified framework for variance component estimation with summary statistics in genome-wide association studies, Ann. Appl. Stat., 11(4): 2027-2051.
https://doi.org/10.1214/17-AOAS1052

Zhu M., and Zhao S., 2007, Candidate gene identification approach: Progress and challenges, Int. J. Biol. Sci., 3(7): 420-427.
https://doi.org/10.7150/ijbs.3.420

International Journal of Molecular Ecology and Conservation, 2026, Vol.16, No.1, 22-30
http://ecoevopublisher.com/index.php/ijmec

Case Study Open Access

Whale-Fall Ecosystems in the Deep Sea Ecological Succession, Biodiversity, and Biogeochemical Significance

Manman Li
Hainan Institute of Biotechnology, Haikou, 570206, Hainan, China
Corresponding author: manman.li@hibio.org

International Journal of Molecular Ecology and Conservation, 2026, Vol.16, No.1 doi: 10.5376/ijmec.2026.16.0003
Received: 10 Jan., 2026
Accepted: 08 Feb., 2026
Published: 28 Feb., 2026

Preferred citation for this article:
Li M.M., 2026, Whale-fall ecosystems in the deep sea ecological succession, biodiversity, and biogeochemical significance, International Journal of Molecular Ecology and Conservation, 16(1): 22-30 (doi: 10.5376/ijmec.2026.16.0003)

Abstract This study systematically analyzes the formation process and ecological succession stages of whale falls, including initial descent, mobile scavenger stage, enrichment opportunist stage, sulfophilic stage, and reef stage, elucidating their roles in maintaining deep-sea biodiversity and nutrient cycling. Key findings highlight that whale falls not only significantly enhance local productivity and species richness but also share ecological and evolutionary links with other deep-sea chemosynthetic environments such as hydrothermal vents and cold seeps, serving as “stepping stones” for species dispersal and adaptive radiation. The crucial functions of whale falls in biogeochemical processes, including carbon sequestration, sulfur cycling, and nitrogen and phosphorus cycling, are also emphasized. However, with the intensification of anthropogenic activities such as whaling, deep-sea mining, bottom trawling, and climate change, the frequency and ecological functions of whale falls are increasingly under threat.

Keywords Whale falls; Deep-sea ecosystems; Biodiversity; Nutrient cycling; Biogeochemistry

1 Introduction

Whale falls, the term for sunken whale carcasses and the ecosystems they create, represent unique and vital habitats in the deep-sea environment. These massive organic inputs deliver concentrated pulses of nutrients to the otherwise food-limited deep ocean, supporting a succession of specialized biological communities and driving significant ecological processes (Smith and Baco, 2003; Smith et al., 2015; Chen and Wang, 2020).

A whale fall is defined as the process and aftermath of a whale carcass descending to the ocean floor, where it forms a localized ecosystem rich in organic matter. These oases provide a substantial energy source for a diverse array of deep-sea organisms, including scavengers, chemosynthetic bacteria, and highly specialized fauna such as bone-eating worms and snails (Smith et al., 2015; Chen and Wang, 2020; Li et al., 2022). Whale falls are considered biodiversity hotspots, supporting thousands of individuals and dozens of species, many of which are new to science or exhibit evolutionary novelties (Smith et al., 2015; Sumida et al., 2016; Chen and Wang, 2020; Li et al., 2022).

The ecological significance of whale falls was first speculated upon in the 19th century, but it was not until the late 20th century that direct observations confirmed their role as unique deep-sea habitats. The discovery in 1989 of a chemoautotrophic community on a whale skeleton in the northeast Pacific marked a turning point, leading to a surge in research and the identification of similar communities in other ocean basins and even in the fossil record dating back 30 million years (Butman et al., 1995; Smith and Baco, 2003; Smith et al., 2015).

Studying whale falls is crucial for understanding deep-sea biodiversity and nutrient cycling. These carcasses act as stepping stones for the dispersal of chemosynthetic organisms, facilitate adaptive radiation, and contribute to the maintenance and connectivity of deep-sea ecosystems (Smith et al., 2015; Sumida et al., 2016; Shimabukuro et al., 2019). The decomposition of whale biomass drives complex successional stages, from scavenger-dominated to chemosynthetic communities, profoundly influencing carbon flux and biogeochemical cycles in the deep ocean (Goffredi et al., 2008; Smith et al., 2015; Chen and Wang, 2020; Amendola et al., 2021; Li et al., 2022).

This study aims to synthesize current knowledge on whale falls, focusing on their ecological roles, successional dynamics, and contributions to deep-sea biodiversity and nutrient cycling. By integrating findings from recent studies, the study seeks to highlight research advances, identify knowledge gaps, and underscore the importance of whale falls as natural laboratories for understanding the functioning and resilience of deep-sea ecosystems.

2 Formation and Stages of a Whale Fall: Succession in Deep-Sea Oases

2.1 Initial fall and physical breakdown

When a whale dies, its carcass sinks rapidly to the ocean floor, often reaching great depths. The descent and initial deposition are influenced by the whale's size, buoyancy, and decomposition gases. Upon arrival, the intact carcass provides a massive, localized input of organic matter to the deep-sea benthos, setting the stage for a series of ecological transformations (Danise et al., 2014; Smith et al., 2014; Bolstad et al., 2023).

2.2 Mobile scavenger stage

The first stage is dominated by large, mobile scavengers such as sharks, hagfish, amphipods, and zoarcid fish. These necrophagous species rapidly consume the soft tissues, often within months to a few years, leaving behind bones and lipid-rich remains. This stage is characterized by intense feeding activity and a sharp increase in local scavenger populations, which can be observed in both modern and fossil whale falls (Danise et al., 2014; Smith et al., 2014; Bolstad et al., 2023; Ibrahim et al., 2024; Serafini et al., 2024).

2.3 Enrichment opportunist stage

As soft tissues are depleted, the whale fall enters the enrichment opportunist stage. Here, the surrounding sediments and exposed bones become colonized by dense populations of opportunistic invertebrates, including polychaete worms (notably Osedax), amphipods, crustaceans, and mollusks. These organisms exploit the remaining organic matter and the enriched sediments, often forming dense assemblages that can persist for months to years (Danise et al., 2014; Smith et al., 2014; Silva et al., 2021; Bolstad et al., 2023; Ibrahim et al., 2024; Serafini et al., 2024).

2.4 Sulfophilic stage

The sulfophilic stage is marked by the anaerobic breakdown of bone lipids, producing hydrogen sulfide. This chemical energy supports chemosynthetic bacteria and a specialized community of symbiotic organisms, such as chemosymbiotic bivalves and sulfur-oxidizing bacteria. The sulfophilic stage can last decades, with the composition and abundance of fauna influenced by the geochemical environment and the amount of remaining organic substrate (Amon et al., 2013; Danise et al., 2014; Onishi et al., 2018; Onishi et al., 2020; Bolstad et al., 2023).

2.5 Reef stage (long-term habitat)

In the final reef stage, after most organic material is exhausted, the remaining bones serve as hard substrate for sessile suspension feeders, including barnacles, bryozoans, corals, and tube-dwelling polychaetes. This stage can persist for years, providing a long-term habitat and contributing to local biodiversity until the bones are buried or fully degraded (Danise et al., 2014; Ibrahim et al., 2024; Serafini et al., 2024).

These stages may overlap, and their duration and community composition can vary with depth, carcass size, and environmental conditions, but together they illustrate the remarkable role of whale falls as dynamic oases in the deep ocean.

3 Ecological Importance of Whale Falls in the Deep Ocean

3.1 Nutrient input in nutrient-poor deep-sea environments

Whale carcasses deliver massive pulses of labile organic matter to the deep-sea floor, a region typically starved of nutrients. This input sustains a succession of scavengers, opportunists, and chemosynthetic organisms, increasing local biomass and altering community structure for years or even decades (Butman et al., 1995; Smith and Baco, 2003; Hilário et al., 2015; Dasgupta et al., 2024). Even smaller mammal falls, such as cow or dolphin carcasses, enrich the seafloor and support diverse assemblages, though their impact is less extensive than that of whale falls

(Hilário et al., 2015; Dasgupta et al., 2024). The organic enrichment from whale falls is a key driver of productivity in otherwise food-limited deep-sea environments (Butman et al., 1995; Smith and Baco, 2003; Dasgupta et al., 2024).

3.2 Enhancement of local biodiversity and creation of ecological hotspots

Whale falls act as biodiversity hotspots, supporting unique assemblages of macrofauna, including generalist scavengers, chemosynthetic fauna, and bone-specialist species. Many species found at whale falls are new to science or rarely observed elsewhere, and the presence of ecosystem engineers like Osedax worms increases habitat complexity and microhabitat diversity (Lucas, 2015; Shimabukuro et al., 2019; Shimabukuro et al., 2022). Studies show that whale-fall communities are distinct from those in surrounding sediments, with higher species richness and evolutionary novelty (Danise et al., 2014; Hilário et al., 2015; Lucas, 2015; Shimabukuro et al., 2019; Shimabukuro et al., 2022). These communities can persist for years, and some taxa exhibit interbasin distributions, highlighting the global significance of whale falls for deep-sea biodiversity (Shimabukuro et al., 2019; Shimabukuro et al., 2022).

3.3 Connection with other chemosynthetic environments

Whale falls share ecological and evolutionary links with other deep-sea chemosynthetic environments, such as hydrothermal vents and cold seeps. Many whale-fall specialists, including chemosymbiotic bivalves and polychaetes, are closely related to or shared with vent and seep communities (Levin et al., 2007; Bernardino et al., 2012; Duperron et al., 2013; Shimabukuro et al., 2019; Avila et al., 2023). Whale falls may serve as “stepping stones” for the dispersal of chemosynthetic fauna, facilitating gene flow and connectivity among spatially isolated habitats (Bernardino et al., 2012; Hilário et al., 2015; Shimabukuro et al., 2019; Avila et al., 2023). The similarity in community structure and reliance on chemosynthetic production underscores the role of whale falls in the broader network of deep-sea reducing ecosystems (Levin et al., 2007; Bernardino et al., 2012; Duperron et al., 2013; Avila et al., 2023).

Whale falls thus represent essential oases in the deep ocean, driving nutrient cycling, supporting high biodiversity, and connecting the patchwork of chemosynthetic habitats across the seafloor.

4 Whale Falls and Biogeochemical Cycles in the Deep Ocean

4.1 Role in carbon sequestration

Whale falls represent a significant mechanism for transferring organic carbon from the surface to the deep ocean. When a whale carcass sinks, it delivers a concentrated pulse of organic carbon to the seafloor. A single large whale can provide an input of organic carbon equivalent to thousands of years of background sedimentation rates (Sheehy et al., 2022). While soft tissues are typically recycled into the food web within about two years, the bones—especially when deposited at depths greater than 1000 meters—can persist for over a century, effectively sequestering carbon and removing it from atmospheric exchange. Restoration of cetacean populations could thus enhance carbon sequestration through increased whale-fall events, contributing to climate change mitigation (Sheehy et al., 2022).

4.2 Influence on sulfur, nitrogen, and phosphorus cycles

Whale falls create localized zones of intense microbial activity, particularly sulfate reduction and methanogenesis, which drive the sulfur cycle in deep-sea sediments (Goffredi et al., 2008; Treude et al., 2009). Sulfate-reducing bacteria break down organic matter, producing hydrogen sulfide that supports chemosynthetic communities similar to those at hydrothermal vents and cold seeps (Treude et al., 2009). Methanogenic archaea also thrive, establishing active methane cycles beneath whale falls (Goffredi et al., 2008). Additionally, cetaceans contribute to nitrogen and phosphorus cycling through the "whale pump"—the vertical and horizontal transport of nutrients via feeding and excretion—which enhances nutrient availability for phytoplankton and supports primary productivity in nutrient-limited waters (Sheehy et al., 2022). This nutrient cycling is crucial for sustaining deep-sea and surface productivity, with estimates suggesting that cetacean-driven processes recycle substantial amounts of nitrogen annually (Sheehy et al., 2022).

4.3 Implications for deep-sea productivity

By delivering organic matter and stimulating chemosynthetic and heterotrophic microbial processes, whale falls enhance local productivity and support complex food webs in the deep sea (Treude et al., 2009; Smith et al., 2015). The enrichment of carbon, sulfur, nitrogen, and phosphorus at whale-fall sites fosters biodiversity and evolutionary innovation, while also linking surface and deep-sea biogeochemical cycles (Goffredi et al., 2008; Treude et al., 2009; Smith et al., 2015; Sheehy et al., 2022). These processes underscore the importance of whale falls as drivers of ecosystem function and productivity in the deep ocean.

5 Specialized Fauna of Whale Falls

5.1 Adaptations of deep-sea species to whale fall habitats

Deep-sea species colonizing whale falls have evolved a suite of adaptations to exploit the rich but ephemeral resources provided by decomposing whale carcasses. Notable adaptations include tolerance to high concentrations of sulfide and other toxic compounds, rapid colonization abilities, and specialized feeding strategies. For example, bone-eating worms of the genus Osedax possess root-like tissues that penetrate bones to extract nutrients, relying on symbiotic bacteria for digestion (Smith et al., 2015; Shimabukuro et al., 2019; Georgieva et al., 2023). Other annelids, such as dorvilleids and hesionids, display trophic niche partitioning and physiological tolerance to the chemically challenging conditions of whale falls, promoting high species diversity and reducing competition (Shimabukuro et al., 2019; Georgieva et al., 2023). These adaptations enable deep-sea fauna to thrive in the unique, resource-rich microhabitats created by whale falls (Smith et al., 2015; Shimabukuro et al., 2019; Georgieva et al., 2023).

5.2 Endemic species and evolutionary implications

Whale falls are hotspots for endemic and newly discovered species. Many taxa found at whale falls, including annelids, mollusks, and crustaceans, are new to science or rarely observed elsewhere (Smith et al., 2015; Sumida et al., 2016; Shimabukuro et al., 2019; Georgieva et al., 2023). The high diversity and endemism, particularly among annelids such as Osedax and Sirsoe, suggest that whale falls have driven adaptive radiation and speciation (Smith et al., 2015; Shimabukuro et al., 2019; Georgieva et al., 2023). Molecular and paleoecological evidence indicates that whale falls have acted as evolutionary stepping stones, facilitating the dispersal and diversification of chemosynthetic fauna between isolated deep-sea habitats like hydrothermal vents and cold seeps (Smith et al., 2015; Sumida et al., 2016; Shimabukuro et al., 2019). This evolutionary connectivity underscores the importance of whale falls in shaping deep-sea biodiversity patterns (Smith et al., 2015; Sumida et al., 2016; Shimabukuro et al., 2019).

5.3 Symbiotic relationships between bacteria and invertebrates

Symbiosis is a defining feature of whale-fall communities. Many invertebrates, such as Osedax worms and bathymodiolin mussels, harbor chemosynthetic bacteria that enable them to utilize the organic and inorganic compounds released during whale decomposition (Lorion et al., 2009; Verna et al., 2010; Smith et al., 2015). Osedax worms, for instance, rely on endosymbiotic bacteria within their root tissues to digest bone-derived organic matter (Verna et al., 2010; Georgieva et al., 2023). Bathymodiolin mussels found on whale falls also maintain specific associations with thioautotrophic bacteria, which are closely related to those found in vent and seep environments (Lorion et al., 2009). These symbiotic relationships are often horizontally transmitted and display high diversity, allowing hosts to adapt to varying substrates and environmental conditions (Lorion et al., 2009; Verna et al., 2010). Such partnerships are central to the success and ecological roles of specialized whale-fall fauna.

6 Human Impacts and Conservation Issues

6.1 Effects of whaling and reduced whale populations on whale fall frequency

Historical and industrial whaling have drastically reduced global whale populations, leading to a significant decline in the frequency of whale falls. This reduction diminishes the input of large organic matter to the deep sea, potentially impacting the unique communities and ecosystem functions that depend on these nutrient-rich oases (Ramirez-Llodra et al., 2011). The loss of whale falls may reduce habitat availability for specialized fauna and disrupt deep-sea nutrient cycling.

6.2 Potential consequences for deep-sea ecosystems

The decline in whale falls can lead to decreased biodiversity and loss of ecosystem services in the deep sea. Whale falls support unique assemblages and contribute to nutrient cycling, so their reduction may have cascading effects on deep-sea food webs and biogeochemical processes (Ramirez-Llodra et al., 2011; Armstrong et al., 2019). Additionally, the deep sea is already facing multiple anthropogenic pressures, including pollution, overfishing, and oil and gas extraction, which further threaten its biodiversity and resilience (Glover and Smith, 2003; Armstrong et al., 2019).

6.3 Deep-sea mining, trawling, and climate change threats

Emerging threats such as deep-sea mining and bottom trawling pose significant risks to deep-sea habitats. Mining activities can cause long-lasting and potentially irreversible damage through habitat destruction, sediment plumes, and pollution, directly impacting both whale falls and the broader deep-sea environment (Levin et al., 2020; Smith et al., 2020; Thompson et al., 2023). Bottom trawling disrupts sedimentary habitats and can reduce carbon sequestration capacity (Levin et al., 2020). Climate change compounds these threats by increasing ocean temperatures, acidification, and hypoxia, which can alter deep-sea community structure and reduce the resilience of ecosystems to other stressors (Ramirez-Llodra et al., 2011; Armstrong et al., 2019; Levin et al., 2020). The cumulative and synergistic effects of these activities are likely to intensify the vulnerability of deep-sea ecosystems, making effective management and conservation strategies increasingly urgent (Ramirez-Llodra et al., 2011; Armstrong et al., 2019; Levin et al., 2020; Smith et al., 2020).

The combined pressures of exploitation, pollution, and climate change highlight the urgent need for comprehensive conservation and management of deep-sea environments, including the protection of whale fall habitats.

7 Case Study: The Monterey Canyon Whale Fall

7.1 Background – discovery and placement of a gray whale carcass

In 2002, a well-preserved gray whale carcass, approximately 9-10 meters long and weighing about 20,000 kg, was discovered at a depth of 2891 meters in Monterey Canyon, California. This site, along with several experimentally implanted carcasses at varying depths, enabled researchers to systematically study whale-fall community development and ecological processes in the deep sea (Goffredi et al., 2004; Lundsten et al., 2010).

7.2 Observations – successional stages documented over several years

Long-term monitoring using remotely operated vehicles (ROVs) revealed that whale-fall communities in Monterey Canyon progress through distinct successional stages. Initial colonization by mobile scavengers (e.g., hagfish, amphipods) was followed by enrichment opportunists and, over time, the establishment of chemosynthetic and bone-specialist fauna. The rate and nature of succession were influenced by depth and environmental conditions, with carcass degradation occurring over sub-decadal timescales (Goffredi et al., 2004; Braby et al., 2007; Lundsten et al., 2010; McGann and Lundsten, 2019).

7.3 Key findings – new species discovery, microbial activity patterns, and food web complexity

The Monterey Canyon whale fall led to the discovery of several new species, including novel polychaetes such as Osedax bone-eating worms, with at least four new species described from the site (Goffredi et al., 2004; Braby et al., 2007). Microbial studies revealed a dynamic succession of methanogenic and sulfate-reducing archaea and bacteria, with methane cycling and elevated carbon concentrations extending up to 10 meters from the carcass (Goffredi et al., 2008; Hasegawa, 2009). The food web was found to be highly complex, involving background deep-sea taxa, opportunists, and chemosynthetic specialists, with Osedax worms acting as foundation species that regulate bone degradation and community succession (Goffredi et al., 2004; Braby et al., 2007; Goffredi et al., 2008; Lundsten et al., 2010).

7.4 Ecological insights – comparisons with other deep-sea chemosynthetic habitats

The Monterey Canyon whale fall shares key features with other chemosynthetic environments, such as hydrothermal vents and cold seeps, including the presence of chemosymbiotic invertebrates and similar microbial

processes (e.g., sulfide and methane production) (Feldman et al., 1998; Braby et al., 2007; Goffredi et al., 2008; Smith et al., 2015). However, the majority of species at Monterey whale falls are background deep-sea taxa, with bone and seep specialists contributing less to overall richness than in some other regions (Lundsten et al., 2010; Smith et al., 2015). The site provides evidence that whale falls can serve as evolutionary stepping stones for vent and seep fauna, supporting the hypothesis of faunal connectivity among deep-sea chemosynthetic habitats (Feldman et al., 1998; Braby et al., 2007; Smith et al., 2015).

The Monterey Canyon whale fall case study highlights the ecological richness, successional dynamics, and evolutionary significance of whale falls as deep-sea oases and their role in connecting chemosynthetic ecosystems.

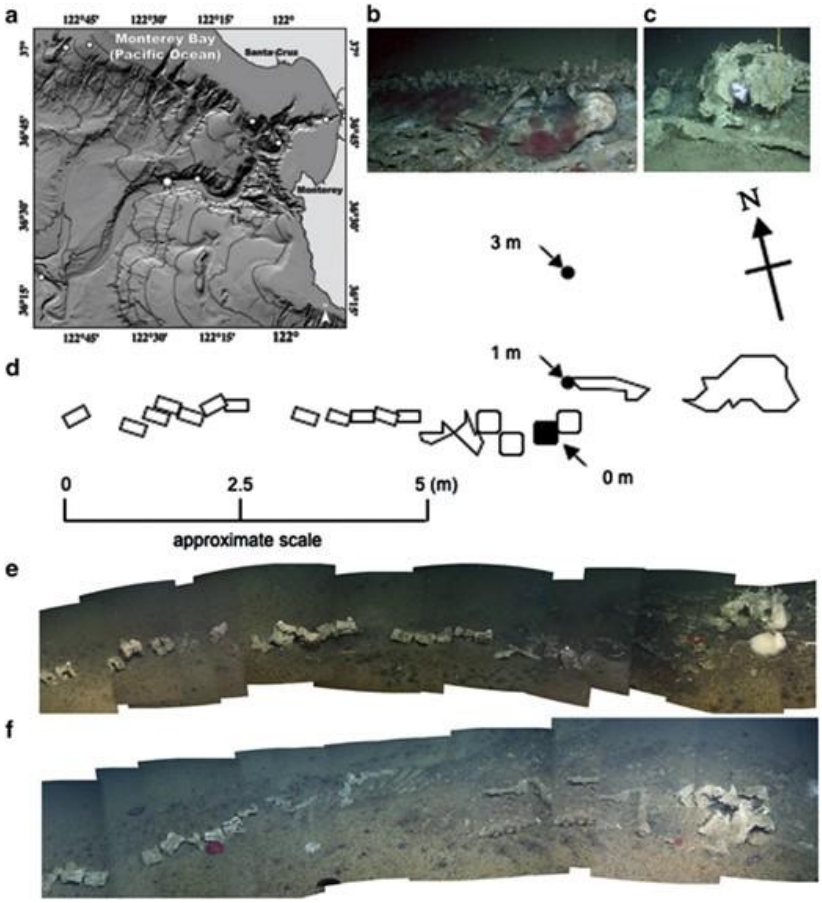

Figure 1(a) Shaded relief map of the continental margin off Monterey Bay showing the whale-fall location at 2891 m depth (large white star). Modified from Goffredi et al., 2004. (b, c, e, f) Photo mosaics of digital still images. (d) Schematic of the whale skeleton at 2893 m showing areas of sediment collections and sampling strategy. (b, c) At 33 months (dive no. T769, November 2004), (e) 45 months (dive no. T917, November 2005) and (f) 51 months (dive no. T991, May 2006) (Adopted from Goffredi et al., 2008)

8 Future Directions in Whale Fall Research

8.1 Gaps in current knowledge

Major gaps remain in the detailed understanding of microbial succession at whale falls. While the broad stages of faunal succession are established, the temporal and spatial dynamics of microbial communities—especially the interplay between sulfate-reducing, methanogenic, and sulfur-oxidizing microbes—are still poorly characterized, partly due to technical challenges in deep-sea sampling and monitoring (Smith et al., 2015; Moriya et al., 2016; Amendola et al., 2021). Additionally, the connectivity between whale falls and other chemosynthetic habitats (e.g., vents, seeps) is not fully resolved. While molecular and faunal evidence suggests whale falls act as evolutionary stepping stones and dispersal corridors for specialized taxa, the extent and mechanisms of this connectivity, particularly for meiofauna and microbes, require further study (Smith et al., 2014; Smith et al., 2015; Avila et al., 2023).

8.2 The role of artificial whale falls in research

Artificial whale falls - experimentally deployed carcasses or large mammal analogs—have become invaluable for studying ecological succession, microbial processes, and faunal colonization in controlled settings (Hilário et al., 2015; Moriya et al., 2016; Aguzzi et al., 2018; Silva et al., 2021). These experiments allow for high-frequency, long-term monitoring and manipulation, helping to overcome the rarity and unpredictability of natural whale falls. Artificial deployments have revealed new species, documented behavioral rhythms, and provided insights into the dispersal and adaptation of deep-sea organisms (Hilário et al., 2015; Aguzzi et al., 2018; Silva et al., 2021). Cow carcasses and whale bones in aquaria or shallow waters have also served as accessible models for testing hypotheses about community assembly and environmental influences (Hilário et al., 2015; Moriya et al., 2016).

8.3 Predicting whale fall distribution with whale migration data

Integrating whale migration and population data with oceanographic models offers a promising avenue for predicting the spatial and temporal distribution of whale falls. Such predictive frameworks could improve estimates of whale fall frequency, guide targeted exploration, and inform conservation strategies by identifying potential biodiversity hotspots and connectivity corridors (Smith et al., 2015). This approach is especially relevant as whale populations recover or shift in response to climate change and human impacts, potentially altering the distribution and ecological role of whale falls in the deep sea (Smith et al., 2015).

Addressing these gaps will deepen understanding of whale falls as dynamic, interconnected oases and their broader significance in deep-sea ecology and evolution.

Acknowledgments

We would like to thank Kris Jin for providing the information and material used in this study.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Aguzzi J., Fanelli E., Ciuffardi T., Schirone A., De Leo F., Doya C., Kawato M., Miyazaki M., Furushima Y., Costa C., and Fujiwara Y., 2018, Faunal activity rhythms influencing early community succession of an implanted whale carcass offshore Sagami Bay, Japan, Scientific Reports, 8. https://doi.org/10.1038/s41598-018-29431-5

Amendola A., Peres F., Moreira J., Sumida P., Paula F., and Pellizari V., 2021, Development of chemosynthetic microbial communities in organic falls deployed in the deep Southwest Atlantic Ocean, bioRxiv. https://doi.org/10.1101/2021.02.16.431415

Amon D., Glover A., Wiklund H., Marsh L., Linse K., Rogers A., and Copley J., 2013, The discovery of a natural whale fall in the Antarctic deep sea, Deep-Sea Research Part II: Topical Studies in Oceanography, 92: 87-96. https://doi.org/10.1016/j.dsr2.2013.01.028

Armstrong C., Vondolia G., Foley N., Henry L., Needham K., and Ressurreição A., 2019, Expert assessment of risks posed by climate change and anthropogenic activities to ecosystem services in the deep North Atlantic, Frontiers in Marine Science. https://doi.org/10.3389/fmars.2019.00158

Avila A., Shimabukuro M., Couto D., Alfaro-Lucas J., Sumida P., and Gallucci F., 2023, Whale falls as chemosynthetic refugia: a perspective from free-living deep-sea nematodes, 10. https://doi.org/10.3389/fmars.2023.1111249

Bernardino Â., Levin L., Thurber A., and Smith C., 2012, Comparative composition, diversity and trophic ecology of sediment macrofauna at vents, seeps and organic falls, PLoS ONE, 7. https://doi.org/10.1371/journal.pone.0033515

Bolstad K., Amsler M., Broyer C., Komoda M., and Iwasaki H., 2023, In-situ observations of an intact natural whale fall in Palmer Deep, Western Antarctic Peninsula, Polar Biology, 46: 123-132. https://doi.org/10.1007/s00300-022-03109-1

Braby C., Rouse G., Johnson S., Jones W., and Vrijenhoek R., 2007, Bathymetric and temporal variation among Osedax boneworms and associated megafauna on whale-falls in Monterey Bay, California, 54: 1773-1791. https://doi.org/10.1016/j.dsr.2007.05.014

Butman C., Carlton J., and Palumbi S., 1995, Whaling effects on deep-sea biodiversity, Conservation Biology, 9: 462-464. https://doi.org/10.1046/j.1523-1739.1995.9020462.x

Chen B., and Wang M., 2020, Whale fall: the creation of a unique marine ecosystem, International Journal of Marine Science. https://doi.org/10.5376/ijms.2020.10.0004

Danise S., Dominici S., Glover A., and Dahlgren T., 2014, Molluscs from a shallow-water whale-fall and their affinities with adjacent benthic communities on the Swedish west coast, Marine Biology Research, 10: 3-16.
https://doi.org/10.1080/17451000.2013.793811

Danise S., Twitchett R., and Matts K., 2014, Ecological succession of a Jurassic shallow-water ichthyosaur fall, Nature Communications, 5.
https://doi.org/10.1038/ncomms5789

Dasgupta S., Lin M., Liu S., Chen S., Xu H., Li J., Ta K., Li S., and Peng X., 2024, Depth and predation regulate consumption of dolphin carcasses in the hadal zone, Deep Sea Research Part I: Oceanographic Research Papers.
https://doi.org/10.1016/j.dsr.2024.104282

Duperron S., Gaudron S., Rodrigues C., Cunha M., Decker C., and Olu K., 2013, An overview of chemosynthetic symbioses in bivalves from the North Atlantic and Mediterranean Sea, Biogeosciences, 10: 3241-3267.
https://doi.org/10.5194/bg-10-3241-2013

Feldman R., Shank T., Black M., Baco A., Smith C., and Vrijenhoek R., 1998, Vestimentiferan on a whale fall, The Biological Bulletin, 194(2): 116-119.
https://doi.org/10.2307/1543041

Georgieva M., Wiklund H., Ramos D., Neal L., Glasby C., and Gunton L., 2023, The annelid community of a natural deep-sea whale fall off eastern Australia, Records of the Australian Museum.
https://doi.org/10.3853/j.2201-4349.75.2023.1800

Glover A., and Smith C., 2003, The deep-sea floor ecosystem: current status and prospects of anthropogenic change by the year 2025, Environmental Conservation, 30: 219-241.
https://doi.org/10.1017/S0376892903000225

Goffredi S., Paull C., Fulton-Bennett K., Hurtado L., and Vrijenhoek R., 2004, Unusual benthic fauna associated with a whale fall in Monterey Canyon, California, Deep-Sea Research Part I: Oceanographic Research Papers, 51: 1295-1306.
https://doi.org/10.1016/j.dsr.2004.05.009

Goffredi S., Wilpiszeski R., Lee R., and Orphan V., 2008, Temporal evolution of methane cycling and phylogenetic diversity of archaea in sediments from a deep-sea whale-fall in Monterey Canyon, California, The ISME Journal, 2: 204-220.
https://doi.org/10.1038/ismej.2007.103

Hasegawa Y., 2009, Characterization of deep-sea microbial activities in a whale-fall sediment.

Hilário A., Cunha M., Génio L., Marçal A., Ravara A., Rodrigues C., and Wiklund H., 2015, First clues on the ecology of whale falls in the deep Atlantic Ocean: results from an experiment using cow carcasses, Marine Ecology, 36: 82-90.
https://doi.org/10.1111/maec.12246

Ibrahim E., Hedeny E., El-Sabbagh A., El-Kheir G., Osman A., Mahdy A., and Ahmed E., 2024, Shallow-water whale-fall communities: evidence from the middle-late Eocene basilosaurid whale bones, Wadi El-Hitan, Fayum, Egypt, Palaeoworld.
https://doi.org/10.1016/j.palwor.2024.06.006

Levin L., Barry J., Felbeck H., Smith C., and Young C., 2007, Advances in vent, seep, whale- and wood-fall biology, Marine Ecology, 28: 1-2.
https://doi.org/10.1111/j.1439-0485.2007.00153.x

Levin L., Wei C., Dunn D., Amon D., Ashford O., Cheung W., Colaço A., Domínguez-Carrió C., Escobar E., Harden-Davies H., Drazen J., Ismail K., Jones D., Johnson D., Le J., Lejzerowicz F., Mitarai S., Morato T., Mulsow S., Snelgrove P., Sweetman A., and Yasuhara M., 2020, Climate change considerations are fundamental to management of deep-sea resource extraction, Global Change Biology, 26: 4664-4678.
https://doi.org/10.1111/gcb.15223

Li Q., Liu Y., Li G., Wang Z., Zheng Z., Sun Y., Lei N., Li Q., and Zhang W., 2022, Review of the impact of whale fall on biodiversity in deep-sea ecosystems, 10.
https://doi.org/10.3389/fevo.2022.885572

Lorion J., Duperron S., Gros O., Cruaud C., and Samadi S., 2009, Several deep-sea mussels and their associated symbionts are able to live both on wood and on whale falls, Proceedings of the Royal Society B: Biological Sciences, 276: 177-185.
https://doi.org/10.1098/rspb.2008.1101

Lucas J., 2015, Faunal composition and community structure of the world's deepest whale-fall community: shedding light on the ecological role of Osedax (Polychaeta: Siboglinidae) and whale-fall biogeography.

Lundsten L., Schlining K., Frasier K., Johnson S., Kuhnz L., Harvey J., Clague G., and Vrijenhoek R., 2010, Time-series analysis of six whale-fall communities in Monterey Canyon, California, USA, 57: 1573-1584.
https://doi.org/10.1016/j.dsr.2010.09.003

McGann M., and Lundsten L., 2019, Influence of an upper bathyal whale-fall on benthic foraminifera of Monterey Canyon, California (USA), Micropaleontology.
https://doi.org/10.47894/mpal.65.5.04

Moriya K., Jenkins R., Sekiguchi T., and Suzuki N., 2016, Succession of whale-fall ecosystems at shallow waters: mainly based on one year monitoring of deployed whale carcasses in aquarium tank and natural environments in Tsukumo Bay, Ishikawa, Japan, Japan Geoscience Union.

Onishi Y., Shimamura S., Yamanaka T., Nakayama R., Ozaki K., Miyazaki M., Tsunogai U., and Fujiwara Y., 2018, Variation of geochemical environments associated with whale-fall biomass mineralization processes in the sediment during the mobile scavenger, enrichment opportunist, and sulfophilic stages, Marine Biology, 165: 1-17.
https://doi.org/10.1007/s00227-018-3398-8

Onishi Y., Yamanaka T., Ozaki K., Nakayama R., Shimamura S., Itami R., Fukushima A., Miyamoto M., and Fujiwara Y., 2020, Relationship between geochemical environments, nutritional resources, and faunal succession in whale-fall ecosystems, Marine Ecology Progress Series, 636: 35-46.
https://doi.org/10.3354/meps13196

Ramirez-Llodra E., Tyler P., Baker M., Bergstad O., Clark M., Escobar E., Levin L., Menot L., Rowden A., Smith C., and Van Dover C., 2011, Man and the last great wilderness: human impact on the deep sea, PLoS ONE, 6.
https://doi.org/10.1371/journal.pone.0022588

Serafini G., Danise S., Maxwell E., Martire L., Amalfitano J., Cobianchi M., Hohenstein T., and Giusberti L., 2024, Of his bones are crinoid made: taphonomy and deadfall ecology of marine reptiles from a pelagic setting (Middle-Upper Jurassic of Northeastern Italy), Rivista Italiana di Paleontologia e Stratigrafia.
https://doi.org/10.54103/2039-4942/22314

Sheehy J., Taylor N., Zwerschke N., Collar M., Morgan V., and Merayo E., 2022, Review of evaluation and valuation methods for cetacean regulation and maintenance ecosystem services with the joint cetacean protocol data, 9.
https://doi.org/10.3389/fmars.2022.872679

Shimabukuro M., Carrerette O., Alfaro-Lucas J., Rizzo A., Halanych K., and Sumida P., 2019, Diversity, distribution and phylogeny of Hesionidae (Annelida) colonizing whale falls: new species of Sirsoe and connections between ocean basins, Frontiers in Marine Science.
https://doi.org/10.3389/fmars.2019.00478

Shimabukuro M., Couto D., Bernardino Â., Souza B., Carrerette O., Pellizari V., and Sumida P., 2022, Whale bone communities in the deep Southwest Atlantic Ocean, Deep Sea Research Part I: Oceanographic Research Papers.
https://doi.org/10.1016/j.dsr.2022.103916

Silva A., Colaço A., Ravara A., Jakobsen J., Jakobsen K., and Cuvelier D., 2021, The first whale fall on the Mid-Atlantic Ridge: monitoring a year of succession, Deep Sea Research Part I: Oceanographic Research Papers.
https://doi.org/10.1016/j.dsr.2021.103662

Smith C., and Baco A., 2003, Ecology of whale falls at the deep-sea floor, Oceanography and Marine Biology, 41: 311-354.

Smith C., Bernardino Â., Baco A., Hannides A., and Altamira I., 2014, Seven-year enrichment: macrofaunal succession in deep-sea sediments around a 30 tonne whale fall in the Northeast Pacific, Marine Ecology Progress Series, 515: 133-149.
https://doi.org/10.3354/meps10955

Smith C., Glover A., Treude T., Higgs N., and Amon D., 2015, Whale-fall ecosystems: recent insights into ecology, paleoecology, and evolution, Annual Review of Marine Science, 7: 571-596.
https://doi.org/10.1146/annurev-marine-010213-135144

Smith C., Tunnicliffe V., Colaço A., Drazen J., Gollner S., Levin L., Mestre N., Metaxas A., Molodtsova T., Morato T., Sweetman A., Washburn T., and Amon D., 2020, Deep-sea misconceptions cause underestimation of seabed-mining impacts, Trends in Ecology and Evolution.
https://doi.org/10.1016/j.tree.2020.07.002

Smith K., Thatje S., Singh H., Amsler M., Vos S., McClintock J., Brown A., Ellis D., Anderson J., and Aronson R., 2014, Discovery of a recent, natural whale fall on the continental slope off Anvers Island, western Antarctic Peninsula, 90: 76-80.
https://doi.org/10.1016/j.dsr.2014.04.013

Sumida P., Alfaro-Lucas J., Shimabukuro M., Kitazato H., Perez J., Soares-Gomes A., Toyofuku T., Lima A., Ara K., and Fujiwara Y., 2016, Deep-sea whale fall fauna from the Atlantic resembles that of the Pacific Ocean, Scientific Reports, 6.
https://doi.org/10.1038/srep22139

Thompson K., Miller K., Wacker J., Derville S., Laing C., Santillo D., and Johnston P., 2023, Urgent assessment needed to evaluate potential impacts on cetaceans from deep seabed mining, 10.
https://doi.org/10.3389/fmars.2023.1095930

Treude T., Smith C., Wenzhöfer F., Carney E., Bernardino Â., Hannides A., Krüger M., and Boetius A., 2009, Biogeochemistry of a deep-sea whale fall: sulfate reduction, sulfide efflux and methanogenesis, Marine Ecology Progress Series, 382: 1-21.
https://doi.org/10.3354/meps07972

Verna C., Ramette A., Wiklund H., Dahlgren T., Glover A., Gaill F., and Dubilier N., 2010, High symbiont diversity in the bone-eating worm Osedax mucofloris from shallow whale-falls in the North Atlantic, Environmental Microbiology, 12(8): 2355-2370.
https://doi.org/10.1111/j.1462-2920.2010.02299.x

International Journal of Molecular Ecology and Conservation, 2026, Vol.16, No.1, 31-43
http://ecoevopublisher.com/index.php/ijmec

Research Report Open Access

Integration Stability and Phenotypic Regulation in Genetically Engineered Livestock from a Molecular Ecology Perspective

Qibin Xu
Animal Science Research Center, Cuixi Academy of Biotechnology, Zhuji, 311800, Zhejiang, China
Corresponding author: Qibin.xu@cuixi.org

International Journal of Molecular Ecology and Conservation, 2026, Vol.16, No.1 doi: 10.5376/ijmec.2026.16.0004
Received: 11 Jan., 2026
Accepted: 09 Feb., 2026
Published: 28 Feb., 2026

Preferred citation for this article:
Xu Q.B., 2026, Integration stability and phenotypic regulation in genetically engineered livestock from a molecular ecology perspective, International Journal of Molecular Ecology and Conservation, 16(1): 31-43 (doi: 10.5376/ijmec.2026.16.0004)

Abstract This study systematically analyzes the genetic stability and phenotypic evaluation framework of transgenic livestock, providing a comprehensive overview of key technologies for phenotypic data collection and analysis. It focuses on the integration patterns of exogenous genes within host genomes, the effects of copy number variation, and the stability of gene expression across generations. Through case studies on transgenic cattle and pigs, the research reveals a strong correlation between genetic stability and phenotypic consistency. The results indicate that combining non-viral vectors with gene-editing technologies not only facilitates long-term stable gene expression but also effectively ensures physiological health and production reliability in livestock. This study offers a solid theoretical foundation and methodological support for the safety assessment and industrial application of transgenic livestock.

Keywords Transgenic livestock; Genetic stability; Phenotypic evaluation; Gene editing; Multi-omics integration

1 Introduction

With the rapid advancement of molecular biology and genetic engineering technologies, transgenic technology has become an essential tool in modern livestock breeding. It offers unprecedented opportunities to improve production performance, disease resistance, and the quality of animal products (Wheeler and Walters, 2001; Laible et al., 2015; Hryhorowicz et al., 2020). By introducing exogenous genes into animal genomes, scientists can endow livestock with new economic traits, such as accelerated growth rate, improved feed conversion efficiency, enhanced disease resistance, or superior product quality (Laible et al., 2015; Shaukat, 2021). For instance, transgenic cows can produce milk enriched with specific nutrients, while transgenic pigs are widely used in human disease modeling and xenotransplantation studies (Hryhorowicz et al., 2020; Park, 2023). These groundbreaking achievements have not only transformed livestock production systems but also provided new insights into biomedicine, food safety, and ecological sustainability. However, the research and application of transgenic livestock face critical scientific challenges related to genetic stability and phenotypic consistency, which pose obstacles at the technical, ethical, and regulatory levels (Laible et al., 2015; Eriksson et al., 2018).

Genetic stability serves as the foundation of transgenic livestock research, determining the accuracy and long-term controllability of exogenous gene transmission across individuals and generations (Van Cott et al., 1997; Yum et al., 2018). If the insertion site of an exogenous gene in the chromosome is unstable or if copy number variations and epigenetic modifications occur, this may lead to gene silencing, expression deviation, or undesirable phenotypes, thereby compromising the reliability of research results and the predictability of animal performance (Pursel et al., 1989; Evangelou et al., 2018). For example, studies have shown that certain transgenic pigs initially exhibit the desired disease resistance, but this trait diminishes in subsequent generations, suggesting that the exogenous gene may be influenced by epigenetic regulation or genomic rearrangement (Yum et al., 2018). Similar phenomena have been observed in transgenic sheep and goats, where expression levels of milk protein genes vary among individuals (Van Cott et al., 1997; Evangelou et al., 2018). Therefore, systematic evaluation of genetic stability in transgenic livestock not only helps elucidate the integration behavior of exogenous genes within host genomes but also provides theoretical guidance for improving gene-editing strategies and transformation efficiency (Laible et al., 2015; Wang et al., 2022).

Phenotypic consistency is a crucial indicator for assessing the practical value of transgenic livestock. Even if the insertion and expression of an exogenous gene are relatively stable, inconsistent phenotypic manifestations across different environments, sexes, or genetic backgrounds can undermine both the scientific significance and application potential of the transgenic line (Van Cott et al., 1997; Evangelou et al., 2018). Phenotypic consistency encompasses multiple aspects, including physiological metabolism, reproductive capacity, immune response, and behavioral characteristics. Through systematic phenotypic assessment, researchers can determine whether the introduced gene functions as intended and whether it exerts unintended effects on animal health or growth. For instance, although certain transgenic cows maintain stable gene expression, their milk composition fluctuates abnormally, suggesting that environmental factors and gene network regulation play critical roles in phenotype formation (Yum et al., 2018; Yum et al., 2024). Establishing a standardized phenotypic evaluation system is therefore essential for promoting the transition of transgenic livestock from laboratory research to industrial application (Hryhorowicz et al., 2020; Park, 2023).

This study aims to systematically investigate the methods for evaluating genetic stability and phenotypic performance in transgenic livestock lines, elucidating their intrinsic relationship and implications for breeding practices. By summarizing mainstream detection technologies, analyzing stability variations under different transgenic strategies, and integrating genetic and phenotypic data from representative cases (such as transgenic cattle and pigs), this study seeks to construct a scientific and reproducible evaluation framework. Through this systematic analysis, the study aspires to advance transgenic livestock research toward greater scientific rigor and application sustainability, thereby providing robust support for the modernization of animal husbandry and innovation in life sciences.

2 Transgenic Livestock Technologies

2.1 Common transgenic methods

The construction of transgenic livestock relies on various efficient and stable technologies for introducing and expressing foreign genes. Among them, pronuclear microinjection was the first method to be widely used. This technique involves directly injecting target DNA into the pronucleus of a fertilized egg, allowing the exogenous gene to integrate randomly into the host genome and thereby producing transgenic individuals (Robl et al., 2007).

This approach is relatively simple to perform and applicable to a wide range of species, laying the foundation for the creation of early transgenic animal models such as pigs, cattle, and sheep (Niemann and Kues, 2003). However, due to its uncontrolled integration sites, high mosaicism rates, and significant variability in gene expression, its transgenic efficiency remains low—only about 1%-2%—which limits its suitability for modern commercial breeding applications (Robl et al., 2007).

To overcome these limitations, researchers developed viral vector–mediated transduction. This method employs vectors such as lentiviruses or retroviruses to deliver target genes into the host genome. Viral vectors exhibit a high integration rate and broad host range, allowing for tissue-specific expression of transgenes.

In the past decade, the advent of genome editing technologies has brought revolutionary progress in transgenic livestock production. Among these, the CRISPR/Cas9 system has emerged as the mainstream tool due to its simplicity, low cost, and ability to perform multi-site genome editing. This technique uses an RNA-guided nuclease to induce double-strand breaks at target sites, enabling precise insertion of exogenous genes or knockout of endogenous genes through homology-directed repair (HDR) (Table 1).

Additionally, somatic cell nuclear transfer (SCNT) is often used in combination with the CRISPR system to clone embryos that have been successfully edited (Robl et al., 2007). This integration of techniques significantly improves transgenic efficiency and provides a more reliable model for studying phenotypic stability and genetic consistency in transgenic livestock.

Table 1 Comparison of different transgenic vector methods

Method	Efficiency	Control of Integration Site	Safety Risk	Typical Applications
Microinjection	Low	Random	Risk of uncontrolled insertion sites	Early models of mice, pigs, and cattle
Viral Vector	Medium	Random but stable integration	Risk of insertional mutagenesis	Transgenic large animals
Gene Editing (CRISPR)	High	Site-specific	Controllable off-target risk	Disease-resistant pigs, transgenic cattle, chickens

2.2 Major application areas

Transgenic livestock technology demonstrates broad and profound application potential in both agriculture and biomedicine. In growth performance improvement, the introduction of genes regulating growth hormone (GH) or insulin-like growth factor (IGF) can significantly enhance animal growth rate and feed conversion efficiency (Niemann and Kues, 2003). In disease resistance enhancement, researchers have successfully developed livestock resistant to specific pathogens by introducing antiviral or immune-related genes. For instance, transgenic cattle carrying the NRAMP1 gene exhibit remarkable resistance to tuberculosis, while pigs expressing interferon show improved defense against viral infections compared with conventional breeds.

In terms of product quality optimization, transgenic techniques can regulate metabolic pathways involving milk proteins, fatty acids, and amino acids, enabling the production of dairy and meat products with higher nutritional value and enhanced functionality (Wheeler and Walters, 2001).

Moreover, transgenic livestock are widely used as animal bioreactors for the efficient production of pharmaceutical proteins. Cows and goats capable of expressing recombinant human antithrombin III, insulin, or human serum albumin in their milk have become essential platforms for the continuous production of high-value biopharmaceuticals (Bertolini et al., 2016).

2.3 Ethical and regulatory challenges arising from technological development

Despite the significant scientific and economic value brought by transgenic livestock technology, its development continues to raise complex ethical debates and regulatory challenges.

On the ethical level, public concerns focus on issues such as animal welfare, the naturalness of life, and potential health risks (Eriksson et al., 2018). Early studies revealed that some transgenic animals exhibited reduced fertility and immune system disorders during experiments, sparking ethical reflections on the boundaries of biotechnological intervention in living organisms.

On the regulatory level, policy orientations differ significantly across countries. The United States and Europe typically adopt risk-based regulatory frameworks, emphasizing the properties and safety of the final product; whereas Asian regions tend to prioritize process-based supervision and ethical review (Bertolini et al., 2016). In China, research and commercialization of transgenic livestock are approached with prudence, following the guiding principle of “scientific research first, regulation in parallel”, which emphasizes both technological innovation and strict oversight.

3 Theoretical Basis of Genetic Stability

3.1 Definition and importance of genetic stability

Genetic stability refers to the ability of an exogenous gene to maintain structural integrity, a relatively constant level of expression, and stable transmission to offspring according to genetic laws across generations within the host genome. Its connotation encompasses three aspects: structural stability (absence of unintended mutations, rearrangements, or deletions), expression stability (absence of sustained silencing or abnormal fluctuations), and transmission stability (following Mendelian segregation and predictable inheritance in populations) (Yum et al., 2018; Yum et al., 2024).

For transgenic livestock, genetic stability not only determines the reproducibility and controllability of experimental outcomes but also directly affects the feasibility and safety of breeding and industrial application (Yum et al., 2018; Yum et al., 2024). Long-term population tracking has indicated that livestock produced via non-viral integration strategies such as transposon systems exhibit no significant accumulation of somatic mutations, abnormal copy number variations, or telomere anomalies after years of breeding and multigenerational propagation. This suggests that such methods maintain high levels of genomic integrity and physiological health (Yum et al., 2018; Yum et al., 2024).

3.2 Effects of insertion site and copy number on stability

The insertion site and copy number are critical determinants of genetic stability and predictable gene expression (Table 2). Random integration methods—such as pronuclear microinjection or certain viral vectors—are prone to position effects. When exogenous genes are inserted into heterochromatic regions, repetitive sequences, or areas near active transposable elements, they often experience epigenetic silencing, expression drift, or integration rearrangements. In contrast, integration into genomic safe harbors can substantially reduce insertional mutagenesis risks and enhance expression stability.

Regarding copy number, while high copy numbers may initially increase expression levels, they also elevate the likelihood of homologous recombination or tandem repeat-induced instability and silencing. Conversely, low copy numbers, particularly single-copy targeted insertions, favor long-term stable expression and predictable inheritance. Studies on transgenic goats and cattle have demonstrated that expression levels are not linearly correlated with copy number, indicating that local chromatin environment, promoter selection, and epigenetic modifications are equally crucial in maintaining genetic stability (Yum et al., 2018; Yum et al., 2024).

Table 2 Comparison of the effects of insertion site and copy number on genetic stability and expression

Factor	Favorable Conditions	Unfavorable Conditions	Main Risks	Countermeasures and Recommendations
Insertion Site	Euchromatin regions, genomic safe harbors, distant from key regulatory areas	Heterochromatin, repetitive sequences, regions near active transposons	Position effect, gene silencing, structural variation	Select genomic safe harbors; perform breakpoint sequencing; combine with homologous recombination repair
Copy Number	Single-copy or low-copy targeted integration	High-copy tandem insertions	Recombination, silencing, expression drift	Use single-copy knock-in (KI); perform ddPCR/qPCR quantification; optimize expression control
Regulatory Elements	Species-matched promoters, use of insulators/barriers	Heterologous strong promoters without protective elements	Epigenetic silencing, abnormal histone modifications	Introduce insulators; apply site-specific enhancer strategies

3.3 Molecular mechanisms of exogenous gene inheritance and potential variation risks

The inheritance of exogenous genes in livestock populations follows Mendelian laws, and their stability is influenced by multiple factors such as integration mechanisms, epigenetic regulation, and the host genomic environment. Transposon systems represented by Sleeping Beauty and PiggyBac integrate through a “cut-and-paste” mechanism, tending to insert into non-coding or low-risk regions, thereby reducing the potential hazards of insertional mutagenesis and ensuring the stable transmission of exogenous genes in the germline. Whole-genome sequencing and copy number variation analyses have shown that this type of integration has minimal impact on global genomic stability indicators such as SNP profiles, CNV, and telomere length (Yum et al., 2018).

The potential risks are mainly reflected in three aspects. In terms of structural variation, if the integration site is close to coding regions or key regulatory elements, it may lead to large fragment deletions, rearrangements, or functional disturbances. In terms of epigenetic silencing, changes in promoter methylation or histone modifications may cause intergenerational decreases in expression levels or even gene inactivation. Regarding

positional effect variation, changes in local chromatin plasticity may lead to expression differences among individuals or generations.

To address these issues, researchers have proposed a variety of optimization strategies, including using safe harbor targeting and breakpoint sequencing to ensure integration site safety, adopting single-copy targeted knock-in techniques to reduce structural interference risks, introducing insulator or barrier sequences to prevent epigenetic silencing, selecting promoters that match the host species, and conducting long-term population tracking to verify genetic stability. Long-term follow-up results show that livestock lines established following these design principles generally achieve stable inheritance and reproducible expression of exogenous genes, and no significant negative effects on animal health or genomic integrity have been observed (Yum et al., 2018; Yum et al., 2024).

4 Theoretical Framework and Indicator System for Phenotypic Evaluation

4.1 Significance of phenotypic evaluation in transgenic research

Phenotypic evaluation is the core link connecting genotype, expression, and functional traits, and it is a key method for verifying the biological effects and application value of exogenous genes. The goal of transgenesis is not only to achieve stable integration and expression of exogenous genes but also to continuously and reproducibly achieve the expected trait improvements (such as growth performance, disease resistance, and product quality). Therefore, it is necessary to conduct systematic, quantitative, and long-term phenotypic tracking of individuals and populations to identify expected effects and possible unintended effects (off-target/pleiotropy), and to provide evidence for safety and ethical assessments.

With the development of high-throughput and digital technologies, phenotypic evaluation is evolving from traditional single-trait observation toward multimodal integration (imaging phenomics, metabolomics/proteomics, wearable/environmental sensing, and behavioral monitoring), and combined modeling with genomic information, thereby improving the accuracy of breeding selection and enhancing the depiction of environmental adaptability and animal welfare.

4.2 Common phenotypic indicators: growth rate, physiological metabolism, reproductive performance, disease resistance

The phenotypic indicators of transgenic livestock vary according to target traits and application needs, typically covering five aspects: growth, metabolism, reproduction, health and disease resistance, and welfare and environmental adaptability (König and May, 2019).

For growth rate and body conformation, common indicators include body weight, average daily gain (ADG), feed conversion ratio (FCR), skeletal development, and body composition. For genetic modifications related to the GH/IGF axis and muscle development, the focus should be on verifying energy utilization efficiency and tissue development characteristics (Pinkert, 2014). The introduction of imaging measurement and automatic weighing technologies makes dynamic monitoring more accurate and objective, effectively reducing human error.

In terms of physiological metabolism and endocrinology, energy metabolism indicators such as blood glucose, insulin, cholesterol, and lipid profiles can be monitored, as well as endocrine levels including GH, IGF, and the thyroid axis. Combined with physiological parameters of organs such as liver enzymes and kidney function, and metabolomic and proteomic analyses, these help assess the remodeling of metabolic pathways and potential health risks caused by transgenic manipulation (Pinkert, 2014).

For reproductive performance and offspring traits, indicators such as age at first mating, estrous cycle, semen quality, conception rate, embryo survival and implantation rate, litter size, and lactation ability are examined to assess the impact of exogenous genes on the reproductive axis and their intergenerational inheritance effects.

In terms of disease resistance and immune response, for transgenic modifications aimed at improving disease resistance (such as antiviral, immune regulation, or pathogen recognition receptors), immunological tests

(antibody titer, cytokine profile), pathogen challenge experiments, and transcriptomic analyses can be combined to comprehensively assess infection rate, recovery time, and survival rate (König and May, 2019).

In animal welfare and adaptability, behavioral performance, stress physiological indicators (such as cortisol level), heat stress scores, feeding patterns, and activity rhythms are monitored to prevent excessive pursuit of productivity at the expense of animal welfare, while meeting the ethical concerns of regulators and the public (König and May, 2019).

4.3 Environmental influences on phenotypic expression and their control

The environment is a key external variable determining the reproducibility and generalizability of phenotypic traits. Genotype × Environment interactions (G×E) can cause the same genetic modification to exhibit heterogeneous performance under different ecological and management systems (Figure 1).

Common influencing pathways include: heat and humidity stress altering energy allocation and feed intake; diet composition regulating lipid metabolism and immunity; and density and light exposure affecting reproductive hormones and behavior. To ensure scientific validity and comparability, randomized block or stratified designs should be adopted during the design stage, and continuous records of temperature, humidity, feed composition, pathogen exposure, and stress levels should be maintained. During statistical analysis, covariate correction and stratified analysis should be introduced to ensure scientific rigor and comparability.

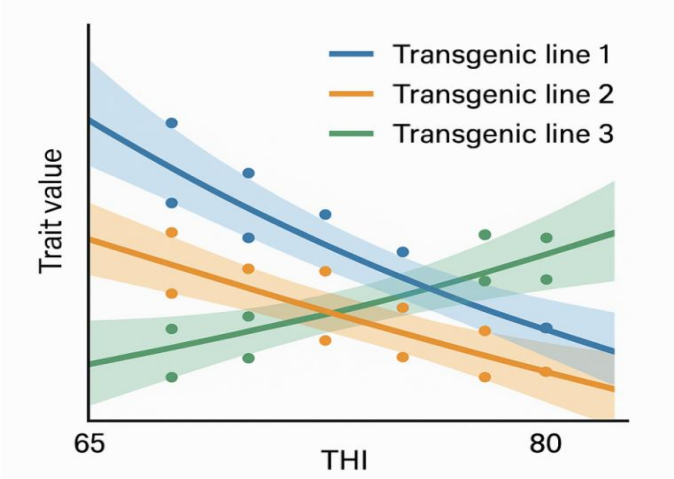

Figure 1 Reaction norms of key traits across THI gradients with random regression fits

5 Phenotypic Evaluation Methods for Transgenic Livestock

5.1 Principles of experimental design: control group setup and sample size

For the setup of control groups, non-transgenic individuals of the same breed, age, and management conditions should be selected as negative controls to minimize interference from genetic background and environmental factors. When studies involve different transgenic constructs, multiple treatment groups should be established, including a vector control, to identify potential nonspecific effects caused by regulatory elements. To further reduce environmental variation, it is recommended to use randomized grouping and block design, maintaining consistency in feeding management, housing density, lighting conditions, and disease prevention measures.

In determining sample size, a balance should be achieved between statistical power and experimental feasibility. Too small a sample size may lead to insufficient statistical power and high variability in results, while excessively large sample sizes may waste resources. For transgenic livestock, especially in large animal experiments, power analysis should be conducted based on the expected effect size, significance level (α), and statistical power (1-β) to determine a reasonable sample size. For high-dimensional studies such as metabolomics, simulation-based sample optimization methods can be used, combined with hybrid sampling strategies that integrate random sampling and extreme phenotype sampling, thereby improving parameter estimation and model prediction accuracy.

5.2 Phenotypic data collection techniques: imaging measurement, metabolomics, and behavioral analysis

Phenotypic data collection is rapidly evolving from traditional manual measurement to high-throughput, automated, non-invasive, and multimodal integration approaches. This transformation has greatly enhanced data accuracy, efficiency, and reproducibility, providing a solid foundation for the systematic phenotypic characterization of transgenic livestock.

In imaging measurement, RGB and depth cameras, laser scanning, structured light, and 3D reconstruction technologies are widely used for body size measurement, weight estimation, and surface or muscle thickness analysis. CT and MRI can analyze tissue distribution and fat deposition patterns. When combined with computer vision algorithms such as convolutional neural networks (CNNs), these technologies enable automatic landmark recognition and phenotypic feature quantification, reducing human error and improving assessment efficiency.

In metabolomics and biochemical detection, gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS) can be used to analyze dynamic changes in energy metabolism, lipid synthesis, and amino acid utilization in samples such as blood, urine, and milk. Combined with transcriptomic and proteomic data, these methods can construct metabolic pathway networks, revealing potential systemic metabolic effects induced by genetic modifications.

In behavioral and physiological monitoring, video tracking systems and wearable sensors enable continuous recording of feeding, rumination, activity rhythms, social behavior, and stress responses. Monitoring heart rate, body temperature, and respiration rate provides quantifiable indicators of livestock adaptability and recovery speed under environmental changes or pathogen exposure, offering multidimensional data support for health and welfare assessment.

5.3 Data analysis and statistical models: multivariate regression, genetic and environmental interaction analysis

The analysis of multidimensional high-throughput data requires an integrated statistical framework to effectively extract genetic signals and reveal the patterns of gene–environment interaction (G×E).

In multivariate analysis, when multiple correlated phenotypes are studied, multivariate regression, canonical discriminant analysis, or stepwise discriminant analysis can be applied for joint modeling. These methods enhance the detection power of quantitative trait loci (QTLs) and gene effects by accounting for trait correlations. For datasets with nonlinear relationships or strong collinearity, partial least squares regression (PLSR) or principal component regression (PCR) can be used for dimensionality reduction before further analysis to reduce noise and improve model stability.

In gene–environment interaction studies, linear mixed models (LMMs) and random regression models (RRMs) are widely used to analyze longitudinal data and multi-environment experimental results. These models describe reaction norms of gene effects across environmental gradients or time, thus revealing genotype performance differences and adaptive traits under varying environments.

In multilevel and machine learning modeling, for complex high-dimensional data such as imaging, metabolomics, and behavioral datasets, algorithms such as random forest (RF), support vector machine (SVM), and deep learning can be applied to capture nonlinear patterns and higher-order interactions. These models enable individual-level trait prediction and cluster analysis. To prevent overfitting, cross-validation and independent external validation sets should be incorporated to evaluate model performance and ensure robustness.

6 Correlation Analysis Between Genetic Stability and Phenotypic Consistency

6.1 Relationship models between stability and phenotypic variation

Genetic stability — the structural integrity and consistent expression of exogenous genes across generations — is the foundation for achieving phenotypic consistency. When copy number and integration sites remain stable over generations and expression variance is low, the intergroup dispersion of target traits decreases significantly; conversely, rearrangements, copy number drift, or epigenetic silencing can amplify phenotypic fluctuations (Van Cott et al., 1997).

At the statistical level, multivariate regression, structural equation modeling (SEM), and the animal model can be used to decompose phenotypic variance into genetic and non-genetic components, jointly assessing indicators of genetic stability (such as integration site consistency, single-copy targeting, and expression variance) and phenotypic uniformity. Cross-species evidence also suggests a significant correlation between the stability of gene expression in founder generations and the conservation of phenotypes in progeny — lineages with more stable expression tend to exhibit lower intergenerational phenotypic diversity.

In animal populations, this relationship often manifests as a negative correlation between phenotypic variance and genotypic consistency. For example, in transgenic pig lines, if the copy number and expression level of an exogenous growth hormone gene vary greatly among individuals, the coefficients of variation (CV) for weight gain rate and fat ratio increase significantly. Conversely, when integration sites are fixed and expression is stable, the phenotypic distribution becomes more concentrated. Furthermore, environmental factors (E) can be incorporated into a three-dimensional relationship model:

$$P = G + E + G \times E$$

where G represents genetic stability factors and G × E represents gene-environment interaction effects.

By analyzing variance components across generations, the contribution of genetic instability to phenotypic inconsistency can be quantified, allowing the prediction of trait performance trends of exogenous genes in breeding populations.

6.2 Effects of different transgenic integration sites on phenotypic consistency

The chromatin environment and local regulatory network of the integration site determine expression robustness and inter-individual consistency. A single, well-defined integration site with single-copy expression usually corresponds to more stable recombinant protein yields and lower variability among individuals (Van Cott et al., 1997). Conversely, random integration often causes position effects. If located near coding regions or key regulatory elements, it may cause insertional mutations or expression drift, thereby increasing phenotypic uncertainty.

By inserting exogenous genes into genomic safe harbors and using insulator sequences, position effects can be mitigated, neighboring chromatin influence reduced, and phenotypic consistency improved.

6.3 Molecular links between gene expression regulation and phenotypic stability

At the transcriptional regulation level, selecting appropriate tissue-specific or host-compatible promoters and optimizing the arrangement of regulatory elements can reduce spatiotemporal fluctuations in gene expression, improving the predictability and stability of transgene expression.

At the epigenetic level, DNA methylation status and histone modification patterns determine how gene expression is maintained over time. During generational transmission, accumulation of methylation may cause gene silencing, while the inclusion of insulator or barrier sequences during construct design can effectively block the influence of unfavorable chromatin environments, thereby delaying or preventing silencing.

At the post-transcriptional and translational regulation level, interactions between miRNAs and mRNAs, as well as RNA-binding protein (RBP)-mediated regulation, affect mRNA stability and protein yield. By optimizing sequences and adjusting codon usage, the risk of transgene sequences being targeted by endogenous miRNAs can be reduced, minimizing expression fluctuations and ensuring stable product output.

Genomic structural variations (CNVs and SVs) also play a key role in phenotypic consistency. These variations alter gene dosage and regulatory network structures, directly influencing key economic traits such as growth rate, muscle development, and reproductive performance. Within regulatory networks, redundant elements such as shadow enhancers can maintain target gene expression when disturbances occur, enhancing overall phenotypic robustness.

By integrating genomics, transcriptomics, and epigenomics data and combining them with eQTL analysis, researchers can systematically identify cis- and trans-regulatory loci contributing to expression stability and phenotypic consistency. This multi-omics approach provides molecular-level scientific evidence for screening genetically stable lineages and evaluating the long-term reliability of transgenic livestock.

7 Case Analysis: Genetic Stability and Phenotypic Studies in Transgenic Cattle and Pigs

7.1 Case background: representative experiments and commercial transgenic livestock projects

Cattle and pigs hold significant value in agriculture and biomedicine and are among the most representative species in transgenic livestock research. Transgenic cattle research focuses on optimizing milk composition and expressing exogenous pharmaceutical proteins, while transgenic pig research covers disease-resistant breeding, pharmaceutical protein bioreactors, and xenotransplantation donor construction (Van Cott et al., 1997; Yum et al., 2018; Yum et al., 2024). Over the past decade, transposon-mediated nonviral transfer and targeted gene editing/integration have provided feasible routes to achieve long-term genetic stability and predictable phenotypes (Yum et al., 2018; Yum et al., 2024).

7.2 Genetic stability testing and result analysis

7.2.1 Transgenic cattle (example: transposon-mediated integration and mammary gland expression)

Long-term follow-up has shown that over more than 10 years, physiological indicators and nutritional composition in these cattle show no significant difference compared with control groups. Whole-genome resequencing also revealed no increase in somatic mutation rate, copy number variation (CNV), or structural variation (SV), indicating good maintenance of genomic integrity. The exogenous gene was stably transmitted through the germline, and expression levels remained consistent across generations (Yum et al., 2018; Yum et al., 2024).

7.2.2 Transgenic pigs (two categories: pharmaceutical protein expression and disease resistance gene editing)

Lactating transgenic pigs expressing recombinant human protein C (rhPC) showed a strong correlation between a single integration site plus stable copy number and stable rhPC production in milk. The Western blot isoform profiles were highly consistent within the same lineage, suggesting that post-translational modifications contribute to inter-lineage differences (Van Cott et al., 1997).

For xenotransplantation applications, multi-gene–modified Yucatan miniature pigs achieved knockout of immunogenic loci and expression of human regulatory proteins, showing reduced immunogenicity when co-cultured with human immune cells.

Compared with random integration, site-specific editing (e.g., PRRSV-resistant pigs with targeted CD163 modification) showed no evidence of “reversion mutations” or “structural rearrangements” during generational tracking, and transcriptional balance in nearby regions remained intact, demonstrating that editing specificity and stability are superior to random integration (based on the key points of the second dataset).

Table 3 Evidence matrix of genetic stability

Species/Line	Transgenic Strategy	Integration/Edit Characteristics	Intergenerational Transmission	Whole-Genome Integrity	Key Conclusions
Cattle	Transposon-mediated (non-viral)	Single/few copies, noncoding region integration	expression across F1–F(_n) generations	Mutation rate/CNV/SV ≈ control	Long-term safety and stability (Yum 2018/2024)
Pig (rhPC)	Random integration (early stage)	Single defined site preferable	Stable across multiple litters and generations	Consistent isoform profiles within line	Single-site stability → predictable yield (Van Cott 1997)
Xenotransplantation pigs	Multigene modification	Immunogenic loci knockout + human gene insertion	Validated across generations	In vitro immune compatibility ↑	Significantly reduced immunogenicity

7.3 Phenotypic evaluation and comparison of production performance

7.3.1 Cattle (mammary expression of pharmaceutical proteins / functional improvement of dairy products)

In long-term breeding, growth, reproduction, and health indicators were comparable to controls; no significant differences were found in the nutritional composition of milk and meat, indicating agricultural application potential and food safety (Yum et al., 2024). The target proteins (such as lactoferrin or recombinant human proteins) were highly and stably expressed in milk, and metabolomic/proteomic and functional assays showed enhanced antibacterial and bioactive properties, consistent with the evidence chain presented in Section 7.2.

7.3.2 Pigs (rhPC bioreactor / disease resistance / xenotransplantation)

The milk expression level of rhPC in transgenic pigs ranged from 100-1800 μg/ml among different constructs or lines and remained stable within the same line. Reproductive and lactation performance were normal, making them suitable for large-scale biopharmaceutical production (Van Cott et al., 1997). For xenotransplantation purposes, multi-gene-modified miniature pigs exhibited stable health and physiological performance while successfully expressing multiple human complement-regulating and anticoagulant proteins, significantly improving organ compatibility indicators. In addition, the breed of recipient sows and embryo transfer conditions significantly affected pregnancy and delivery rates, providing a basis for optimizing breeding procedures.

7.4 Research outcomes and risk assessment conclusions

The genetic stability of exogenous genes is the core prerequisite for ensuring phenotypic consistency and production predictability. Defined insertion sites, single-copy integration, and stable expression patterns constitute the foundation of a reliable genetic system. The use of nonviral-mediated gene transfer and targeted integration technologies can significantly reduce the risks of structural variations and epigenetic silencing, thereby enhancing genetic safety and long-term expression consistency (Van Cott et al., 1997; Yum et al., 2024).

In terms of risks and countermeasures, early random integration strategies were often associated with potential issues such as repetitive sequence insertions and expression drift, leading to genetic instability or functional loss. To prevent such risks, comprehensive molecular characterization analyses should be performed, including breakpoint sequencing or whole-genome sequencing (WGS), copy number quantification, and methylation profiling, to accurately assess integration features. Meanwhile, combining insulators and safe harbor site strategies can effectively isolate adverse chromatin effects and reduce the probability of transgenerational silencing. Long-term population tracking can further verify the stable transmission of both genetic and phenotypic traits. Additionally, optimizing recipient breed selection and embryo manipulation procedures in breeding and transplantation stages helps improve reproductive success rates and maintain trait consistency.

At the application and translational level, research and industrial practice in cattle and pigs have fully demonstrated that transgenic livestock hold sustainable potential for agricultural production and biomedicine. The establishment of standardized phenotypic evaluation and continuous safety monitoring systems not only ensures product safety and functional reliability but also provides scientific and transparent evidence for regulatory review and public communication, thereby promoting the social acceptance and regulated development of transgenic technology (Yum et al., 2024).

8 Challenges and Future Directions

8.1 Technical aspects: precision of gene editing and controllability of insertion sites

Although transgenic livestock technologies have made remarkable progress, many challenges remain at the technical level. The precision of gene editing is a primary concern. Current mainstream systems such as CRISPR/Cas9, TALEN, and ZFN possess high editing efficiency but still may cause off-target effects, leading to unintended gene mutations or chromosomal rearrangements. These molecular events may disrupt key genes or regulatory elements, resulting in physiological abnormalities or phenotypic drift, which could affect the reliability of research conclusions and industrial safety. For instance, in some pig-editing experiments, off-target mutations disrupted immune gene expression, leading to reduced disease resistance. Therefore, improving editing system specificity and developing controllable gene repair mechanisms will be central to future technical optimization.

Meanwhile, the controllability of insertion sites directly determines the stability and consistency of exogenous gene expression. The "position effect" caused by traditional random integration remains a major source of phenotypic variation. Although the discovery of "safe harbor" loci (e.g., Rosa26, H11) has greatly improved this issue, safe sites in different species have yet to be systematically identified, and their tissue specificity and transcriptional activity still require further validation. In the future, site-specific targeting systems based on genome editing and precise recombination are expected to become a research focus. By combining high-fidelity Cas variants (such as Cas12, Cas13) with recombinase systems, precise chromosomal integration of exogenous genes at predetermined locations can be achieved, fundamentally eliminating random insertion risks. Furthermore, the use of long-read sequencing and whole-genome validation technologies will enable real-time verification and risk screening of editing results, establishing a safer and more traceable construction process for transgenic livestock.

8.2 Management and ethics: animal welfare and public acceptance

Beyond technical issues, transgenic livestock development faces complex management and ethical challenges. Animal welfare remains a central social concern. While gene editing improves traits, it may also cause physiological burdens and health risks such as metabolic overload, reproductive disorders, or shortened lifespan. For example, certain pigs with high growth hormone expression exhibited myocardial hypertrophy and endocrine imbalance, prompting ethical reflections within the scientific community on the moral boundaries of human intervention in nature. Therefore, in research and breeding practices, animal welfare assessment standards should be established, and the physiological conditions of edited animals should be monitored long-term to ensure that improved traits are not achieved at the cost of animal health.

Public acceptance and regulatory frameworks directly influence the future of transgenic livestock applications. Due to differing public perceptions of "genetically modified animal products", some regions maintain a cautious or even resistant stance toward commercialization. Western countries have implemented limited market access through comprehensive approval and labeling systems, while Asian nations remain in the exploratory phase of regulatory development. In recent years, China has strengthened legal regulation and ethical review of transgenic technology, though improvements in research transparency, public science communication, and risk dialogue are still needed. Enhancing public understanding of genetic science and promoting openness and consistency in regulatory systems are key to fostering rational social consensus and effective policy implementation.

From a global governance perspective, ethical issues surrounding transgenic livestock involve not only animal rights but also biodiversity conservation and food safety. In the future, an international framework for ethical evaluation and information sharing should be established to promote unified standards among nations in technology assessment, data transparency, and ecological risk evaluation, balancing scientific innovation with ethical responsibility.

8.3 Future research trends: multi-omics integration, digital phenotyping, and artificial intelligence evaluation

Future research on transgenic livestock will enter a data-driven and intelligent decision-making era. Multi-omics integration will become the core approach for evaluating genetic stability and phenotypic consistency. By systematically integrating data from genomics, transcriptomics, methylomics, proteomics, and metabolomics, researchers can uncover the dynamic molecular behavior of exogenous genes and their regulatory networks. For instance, combined analysis of DNA methylation and mRNA expression profiles can identify epigenetic regulation of gene silencing; integrating metabolomic and proteomic data can trace the physiological origins of phenotypic variation. This cross-level data fusion will shift research from "gene presence" to "functional realization."

The development of digital phenotyping provides new technological pathways for phenotypic evaluation. Using high-resolution imaging sensors, infrared scanning, behavior recognition, and automated monitoring systems,

researchers can continuously record livestock growth, health, and behavioral patterns. Through cloud databases and data-mining algorithms, precise dynamic phenotypic models can be built to enable individual tracking and group performance prediction. This digital monitoring approach not only improves the objectivity and timeliness of data collection but also supports large-scale breeding and health management with intelligent tools.

Artificial intelligence (AI) and machine learning algorithms will play an increasingly important role in genetic and phenotypic evaluation. AI can automatically identify key factors affecting genetic stability and trait expression from large-scale omics and phenotypic datasets, constructing predictive models. For example, deep neural networks can perform pattern recognition on phenotypic outcomes associated with different editing sites, assisting in selecting optimal insertion sites and gene constructs. In the future, AI will be deeply integrated with bioinformatics platforms to form a Smart Breeding System, enabling integrated decision support for gene design, phenotypic evaluation, environmental monitoring, and risk prediction.

Overall, future research on transgenic livestock will move toward precision, safety, intelligence, and sustainability. Through the coordinated advancement of technological innovation, ethical governance, and intelligent evaluation, transgenic livestock are expected to make forward-looking contributions to global food security, medical health, and agricultural modernization, while ensuring animal welfare and ecological safety.

References

Bertolini L.R., Meade H., Lazzarotto C.R., Martins L.T., Tavares K.C., Bertolini M., and Murray J.D., 2016, The transgenic animal platform for biopharmaceutical production, Transgenic Research, 25(3): 329-343.

Eriksson S., Jonas E., Rydhmer L., and Röcklinsberg H., 2018, Invited review: breeding and ethical perspectives on genetically modified and genome-edited cattle, Journal of Dairy Science, 101(1): 1-17.
https://doi.org/10.3168/jds.2017-12962

Evangelou A., Ignatiou A., Antoniou C., Kalanidou S., Chatzimatthaiou S., Shianiou G., Ellina S., Athanasiou R., Panagi M., Apidianakis Y., and Pitsouli C., 2018, Unpredictable effects of the genetic background of transgenic lines in physiological quantitative traits, G3: Genes|Genomes|Genetics, 9: 3877-3890.
https://doi.org/10.1534/g3.119.400715

Hryhorowicz M., Lipiński D., Hryhorowicz S., Nowak-Terpiłowska A., Ryczek N., and Zeyland J., 2020, Application of genetically engineered pigs in biomedical research, Genes, 11(6): 670.
https://doi.org/10.3390/genes11060670

König S., and May K., 2019, Invited review: phenotypic strategies and quantitative-genetic background of resistance, tolerance, and resilience-associated traits in dairy cattle, Animal, 13(5): 897-908.

Laible G., Wei J., and Wagner S., 2015, Improving livestock for agriculture – technological progress from random transgenesis to precision genome editing heralds a new era, Biotechnology Journal, 10(1): 109-120.
https://doi.org/10.1002/biot.201400193

Niemann H., and Kues W.A., 2003, Application of transgenesis in livestock for agriculture and biomedicine, Animal Reproduction Science, 79(3-4): 291-317.

Park T., 2023, Gene-editing techniques and their applications in livestock and beyond, Animal Bioscience, 36(2): 333-338.
https://doi.org/10.5713/ab.22.0383

Pinkert C.A., 2014, Transgenic animal technology: a laboratory handbook, Newnes, Oxford, pp. 1-691.

Pursel V., Pinkert C., Miller K., Bolt D., Campbell R., Palmiter R., Brinster R., and Hammer R., 1989, Genetic engineering of livestock, Science, 244(4910): 1281-1288.
https://doi.org/10.1126/science.2499927

Robl J.M., Wang Z., Kasinathan P., and Kuroiwa Y., 2007, Transgenic animal production and animal biotechnology, Theriogenology, 67(1): 127-133.

Shaukat M., 2021, Review on transgenic technology in livestock: current status and future horizons, Pakistan Journal of Science, 73(1): 1.
https://doi.org/10.57041/pjs.v73i1.657

Van Cott K., Luboń H., Russell C., Butler S., Gwazdauskas F., Knight J., Drohan W., and Velander W., 1997, Phenotypic and genotypic stability of multiple lines of transgenic pigs expressing recombinant human protein C, Transgenic Research, 6(3): 203-212.
https://doi.org/10.1023/a:1018442124584

Wang S., Qu Z., Huang Q., Zhang J., Lin S., Yang Y., Meng F., Li J., and Zhang K., 2022, Application of gene editing technology in resistance breeding of livestock, Life, 12(7): 1070.
https://doi.org/10.3390/life12071070

Wheeler M.B., and Walters E.M., 2001, Transgenic technology and applications in swine, Theriogenology, 56(8): 1345-1369.
https://doi.org/10.1016/s0093-691x(01)00635-5

Yum S., Choi B., Gim G., Eom K., Lee S., Kim D., Lim E., Kim D., Heo S., Shim D., and Jang G., 2024, Long-term (10-year) monitoring of transposon-mediated transgenic cattle, Transgenic Research, 33(5): 503-512.
https://doi.org/10.1007/s11248-024-00401-0

Yum S., Lee S., Park S., Shin I., Hahn S., Choi W., Kim H., Kim H., Bae S., Lee J., Moon J., Lee W., Lee J., Lee C., Kim S., and Jang G., 2018, Long-term health and germline transmission in transgenic cattle following transposon-mediated gene transfer, BMC Genomics, 19(1): 387.
https://doi.org/10.1186/s12864-018-4760-4

International Journal of Molecular Ecology and Conservation, 2026, Vol.16, No.1, 44-51
http://ecoevopublisher.com/index.php/ijmec

Feature Review Open Access

Epigenetic Regulation of Drought Adaptation in Wild Grass Species

Jiong Fu
Hainan Provincial Key Laboratory for Crop Molecular Breeding, Sanya, 572025, Hainan, China
Corresponding author: Jiong.fu@hitar.org

International Journal of Molecular Ecology and Conservation, 2026, Vol.16, No.1 doi: 10.5376/ijmec.2026.16.0005
Received: 13 Jan., 2026
Accepted: 10 Feb., 2026
Published: 28 Feb., 2026

Preferred citation for this article:
Fu J., 2026, Epigenetic regulation of drought adaptation in wild grass species, International Journal of Molecular Ecology and Conservation, 16(1): 44-51 (doi: 10.5376/ijmec.2026.16.0005)

Abstract This study synthesizes current knowledge on the epigenetic regulation of drought adaptation in wild grass species, with emphasis on DNA methylation, RNA-directed DNA methylation, histone modifications, chromatin accessibility, non-coding RNAs, and stress memory. We argue that epigenetic regulation should be treated not as an isolated molecular layer, but as part of a molecular ecology framework in which repeat-rich grass genomes, local climatic heterogeneity, transposable element control, developmental state, and population history jointly shape drought-responsive phenotypes. Mechanistically, the strongest conserved themes are: maintenance and remodeling of CG, CHG, and CHH methylation by MET1, CMT/SUVH, and RdDM pathways; dynamic coupling between DNA methylation and heterochromatin marks such as H3K9 methylation; the involvement of active chromatin states and accessibility changes in rapid stress-responsive transcription; and the potential for within-generation and, in some cases, transgenerational stress memory. From a molecular ecology perspective, we propose that the next generation of studies in wild grasses should combine environmental gradient sampling, common gardens, reciprocal transplants, and multi-omics assays such as whole-genome bisulfite sequencing, ATAC-seq, ChIP-seq or CUT&Tag, RNA-seq, and small RNA-seq. We then develop a detailed case analysis for Setaria viridis as an ideal wild-grass model for drought epigenomics and outline a publication-ready workflow integrating WGBS, ATAC-seq, histone profiling, and transcriptomics. Finally, we discuss how epigenetic knowledge can inform conservation genomics, restoration, epibreeding, and targeted epigenome editing, while also emphasizing key limitations, including causality, tissue heterogeneity, epigenetic resetting, and the still-limited number of direct field-based studies in natural wild-grass populations. Together, the evidence supports a transition from correlative stress epigenetics to predictive eco-epigenomics for dryland conservation and climate-resilient grass improvement.

Keywords Drought adaptation; Wild grasses; Eco-epigenomics; DNA methylation; RNA-directed DNA methylation; Chromatin accessibility; Stress memory; Setaria viridis; Conservation genomics; Epibreeding

1 Introduction

Grasslands and rangelands are under escalating pressure from climate change, drought, land degradation, and woody encroachment. Recent global assessments have emphasized that large fractions of the world's rangelands are already degraded, and climate-driven drought is increasingly interacting with land-use change to alter ecosystem structure, productivity, hydrology, and biodiversity. Because grasses dominate many open biomes and frequently determine both forage production and belowground carbon dynamics, understanding the biological basis of drought resilience in wild grass species is a conservation problem as much as a plant biology problem (Fortes and Gallusci, 2017). Furthermore, the increasing frequency and intensity of extreme drought events are expected to impose strong selective pressures on natural plant populations, making the identification of adaptive mechanisms a priority for both ecosystem conservation and climate-resilient agriculture (Auge et al., 2023).

At the same time, epigenetics has become central to plant environmental biology because it offers a mechanistic framework through which plants translate external stress signals into altered patterns of gene regulation without changing their underlying DNA sequence. Reviews published over the past decade have converged on the view that plant environmental memory can involve multiple interconnected mechanisms, including DNA methylation, histone modifications, chromatin remodeling, RNA-mediated regulation, and persistent stress priming across developmental stages or even generations (Law and Jacobsen, 2010; Du et al., 2015; Crisp et al., 2016; Lämke and Bäurle, 2017; Auge et al., 2023). In plants, these mechanisms are particularly important because sessile organisms cannot escape unfavorable environmental conditions and must continuously adjust their physiology,

growth, and reproductive strategies in response to environmental fluctuations. Epigenetic regulation therefore provides a potentially rapid and reversible mechanism for enhancing phenotypic plasticity and environmental responsiveness (Gutzat and Mittelsten Scheid, 2012; Springer and Schmitz, 2017).

Wild grasses are especially valuable systems for investigating drought-associated epigenetic adaptation for several reasons. First, many species occupy extensive environmental gradients characterized by substantial variation in precipitation, temperature, and soil moisture availability, resulting in pronounced local adaptation and ecological differentiation. Second, grass genomes are frequently enriched with transposable elements and repetitive DNA, making DNA methylation-mediated genome stabilization and transposon silencing particularly important components of stress adaptation (Sigman and Slotkin, 2016; Wicker et al., 2018). Because environmental stress can alter the activity of transposable elements and the chromatin landscape surrounding them, epigenetic regulation may play a disproportionately important role in grass genome responses to drought (Cavrak et al., 2014; Ito et al., 2011). Third, several wild and semi-wild grasses represent the evolutionary relatives of globally important cereal crops, providing opportunities to translate discoveries from natural systems into crop improvement strategies aimed at enhancing drought tolerance (Springer and Schmitz, 2017). Finally, a limited number of grass species, particularly Setaria viridis, offer the experimental tractability necessary to integrate ecological sampling, functional genomics, and epigenetic analyses within a single framework. As a rapidly cycling, transformable C4 grass with extensive genomic resources, S. viridis has emerged as a powerful model for elucidating the molecular and epigenetic mechanisms underlying drought adaptation in natural grass populations (Brutnell et al., 2010; Jiang et al., 2013; Sebastian et al., 2014).

This study is therefore organized around a simple proposition: drought adaptation in wild grasses emerges from the interaction of ecological selection, population history, genome architecture, and multi-layered epigenetic regulation. The manuscript emphasizes recent literature, uses a molecular ecology lens throughout, and closes with a detailed Setaria viridis case analysis and a forward-looking agenda for conservation, adaptive restoration, and epibreeding.

2 Epigenetic Architecture and Molecular Ecology Framework

2.1 DNA methylation and transposable element regulation in drought adaptation

DNA methylation represents the most extensively studied epigenetic mechanism involved in plant responses to drought stress. In plants, cytosine methylation occurs in three sequence contexts, namely CG, CHG, and CHH, which are maintained through distinct but interconnected pathways. CG methylation is primarily maintained by METHYLTRANSFERASE 1 (MET1), whereas CHG methylation is regulated through the coordinated activities of CHROMOMETHYLASE 3 (CMT3) and SUVH-mediated feedback loops. CHH methylation is largely established and maintained through CMT2 and the RNA-directed DNA methylation (RdDM) pathway, while active DNA demethylation is mediated by DNA glycosylase-dependent mechanisms (Law and Jacobsen, 2010; Du et al., 2015; Bewick et al., 2017; Parrilla-Doblas et al., 2019).

In wild grass species, DNA methylation serves functions beyond simple promoter regulation. It plays a crucial role in silencing repetitive sequences and transposable elements (TEs), thereby maintaining genome stability under environmental stress conditions (Zemach et al., 2013; Sigman and Slotkin, 2016). Because many grass genomes contain exceptionally high proportions of transposable elements, epigenetic regulation of TEs becomes particularly important during drought adaptation. For example, transposable elements account for more than 80% of the wheat genome and substantially influence genome evolution and gene regulation (Wicker et al., 2018). Environmental stresses may induce TE activation or alter chromatin states surrounding TE-adjacent genes, leading to changes in gene expression patterns (Cavrak et al., 2014; Ito et al., 2011). Consequently, drought adaptation in grasses is closely associated with the interaction between epigenetic regulation and genome architecture, highlighting the importance of methylation-mediated control of repetitive genomic regions.

2.2 Histone modifications and chromatin state remodeling under drought stress

Histone modifications constitute a second major layer of epigenetic regulation involved in plant drought responses. Various histone marks influence chromatin structure and accessibility, thereby affecting the transcriptional activity

of stress-responsive genes. Among these modifications, H3K9 methylation is commonly associated with heterochromatin formation and transcriptional repression, whereas H3K4 trimethylation (H3K4me3) is generally linked to active gene expression. In contrast, H3K27 trimethylation (H3K27me3) functions as a repressive mark involved in developmental regulation and gene silencing (Du et al., 2015; Zhao et al., 2019).

Under drought conditions, dynamic changes in histone modifications contribute to the regulation of genes involved in abscisic acid (ABA) signaling, osmotic adjustment, antioxidant defense, and developmental transitions. Importantly, histone modifications do not function independently but interact extensively with DNA methylation pathways. The reciprocal reinforcement between DNA methylation and H3K9 methylation enables coordinated chromatin remodeling at stress-responsive loci (Du et al., 2015; Enke et al., 2011). Therefore, drought adaptation is increasingly viewed as the outcome of integrated epigenetic reprogramming rather than the action of individual epigenetic marks. Recent evidence further suggests that chromatin remodeling complexes contribute significantly to the maintenance and reconfiguration of epigenetic states during environmental stress responses (Yang et al., 2018).

2.3 Chromatin accessibility and non-coding RNA-mediated regulation

Recent advances in high-throughput sequencing technologies have facilitated comprehensive investigations of chromatin dynamics during plant stress responses. Techniques such as ATAC-seq allow the identification of open chromatin regions, whereas ChIP-seq and CUT&Tag enable genome-wide profiling of histone modifications. These approaches provide valuable insights into the regulatory landscapes that govern gene expression under drought stress.

Chromatin accessibility is particularly relevant because it reflects the readiness of genes to respond to environmental stimuli. While transcriptomic analyses reveal changes in gene expression, chromatin accessibility and histone modification profiles offer information about regulatory potential, transcriptional preparedness, and stress memory (Lämke and Bäurle, 2017; Zhao et al., 2019). Consequently, integrating chromatin-level information with transcriptional data is essential for understanding drought-responsive regulatory networks.

In addition to chromatin remodeling, non-coding RNAs play pivotal roles in drought adaptation. Small RNAs, including siRNAs and miRNAs, participate in gene regulation through the RdDM pathway, guiding sequence-specific DNA methylation and transcriptional silencing (Matzke and Mosher, 2014; Erdmann and Lafontaine Picard, 2020). This pathway contributes not only to transposable element repression and genome stability but also to responses to drought, heat, salinity, nutrient deficiency, and other environmental stresses (Popova et al., 2013; Tricker et al., 2012; Xu et al., 2015). Furthermore, studies have demonstrated that mobile small RNAs can regulate genome-wide methylation patterns and contribute to environmental adaptation (Tamiru et al., 2018). For wild grasses, small RNA populations may therefore represent critical regulators of drought adaptation, providing regulatory information that cannot be detected through conventional protein-coding transcriptome analyses (Zhao and Chen, 2014).

2.4 Stress memory and the molecular ecology framework of drought adaptation

Stress memory has emerged as an important concept in plant environmental adaptation. Plants exposed to a previous drought event often exhibit enhanced responses during subsequent stress episodes, a phenomenon known as drought stress memory. This adaptive response may involve persistent DNA methylation patterns, stable histone modifications, or prolonged chromatin accessibility changes that facilitate faster reactivation of stress-responsive genes (Crisp et al., 2016; Lämke and Bäurle, 2017; Auge et al., 2023).

Although several studies have reported transgenerational inheritance of stress-induced epigenetic states, the stability and ecological significance of such inheritance remain subjects of ongoing debate (Hauser et al., 2011; Blevins et al., 2014; Gutzat and Mittelsten Scheid, 2012). Therefore, drought memory in wild grasses should be considered as a continuum encompassing immediate stress responses, recovery processes, repeated-stress priming, and potential progeny effects rather than as a single phenomenon.

From a molecular ecology perspective, understanding drought adaptation requires moving beyond experiments involving individual genotypes under controlled conditions. Instead, ecologically meaningful investigations should integrate natural population sampling, environmental gradient analyses, common garden experiments, and repeated drought treatments with multi-omics profiling. Such approaches enable researchers to distinguish environmentally induced epigenetic variation from genetically determined differences and to evaluate how epigenetic mechanisms contribute to plant performance and fitness under drought conditions (Fortes and Gallusci, 2017; Springer and Schmitz, 2017). Population epigenomics, methylation-expression association studies, methylQTL analyses, and genotype-by-environment interaction models collectively provide powerful tools for linking epigenetic variation to ecological adaptation and evolutionary processes in wild grass species.

3 Case Analysis in Setaria viridis

3.1 Ecological significance and model value of setaria viridis

Setaria viridis has emerged as one of the most promising model species for investigating epigenetic mechanisms underlying drought adaptation in wild grasses. As a wild C4 grass closely related to several economically important cereal crops, including foxtail millet and other drought-tolerant grasses, S. viridis occupies a unique position at the intersection of ecological adaptation, crop evolution, and molecular genetics. Its short life cycle, relatively small genome, ease of transformation, and well-established crossing protocols make it particularly suitable for integrative molecular ecology studies (Brutnell et al., 2010; Jiang et al., 2013; Sebastian et al., 2014).

Unlike many cultivated crop species that have undergone extensive artificial selection, S. viridis populations retain substantial natural genetic and ecological variation. This diversity provides an excellent opportunity to investigate how wild grass populations adapt to contrasting environmental conditions and whether epigenetic mechanisms contribute to drought resilience. Consequently, S. viridis represents an ideal system for examining the relative contributions of phenotypic plasticity, genetic variation, and epigenetic regulation in shaping adaptive responses to water limitation.

3.2 Experimental framework for investigating drought-induced epigenetic variation

A robust molecular ecology framework for studying drought adaptation in S. viridis should incorporate both ecological sampling and multi-omics analyses. Natural accessions collected across aridity gradients can be grown under controlled conditions and subjected to multiple water regimes, including well-watered controls, progressive drought stress, and repeated drought-recovery cycles. Such experimental designs enable researchers to distinguish immediate stress responses from persistent adaptive changes.

Comprehensive epigenomic analyses should include whole-genome bisulfite sequencing (WGBS) to characterize DNA methylation dynamics, RNA sequencing (RNA-seq) to quantify transcriptional responses, ATAC-seq to assess chromatin accessibility, and ChIP-seq or CUT&Tag approaches to examine histone modifications such as H3K4me3 and H3K27me3 (Law and Jacobsen, 2010; Du et al., 2015; Lämke and Bäurle, 2017). The incorporation of small RNA sequencing would further facilitate the identification of RdDM-associated regulatory pathways involved in drought adaptation (Matzke and Mosher, 2014; Erdmann and Lafontaine Picard, 2020). Physiological traits including leaf water potential, stomatal conductance, abscisic acid accumulation, root architecture, and biomass allocation should also be evaluated to establish links between epigenetic variation and adaptive performance.

3.3 Drought-responsive regulatory networks and epigenetic memory

Current knowledge from plant stress epigenetics suggests that drought adaptation in S. viridis is likely mediated through coordinated changes in multiple regulatory layers. Immediate responses to water deficit are expected to involve rapid transcriptional activation of genes associated with ABA signaling, osmotic adjustment, reactive oxygen species detoxification, and stomatal regulation. These transcriptional responses are often accompanied by increased chromatin accessibility at stress-responsive loci and dynamic alterations in histone modification patterns (Crisp et al., 2016; Lämke and Bäurle, 2017).

DNA methylation changes are predicted to occur more selectively, particularly within transposable element-rich regions and promoter-adjacent regulatory sequences. Repeated drought exposure may induce persistent epigenetic modifications that remain detectable after recovery, thereby contributing to drought stress memory. Such memory-associated loci may exhibit sustained chromatin accessibility or retention of active histone marks, allowing more rapid and robust transcriptional responses during subsequent drought events (Auge et al., 2023). These observations support the hypothesis that epigenetic memory functions as an adaptive mechanism enabling wild grasses to cope with fluctuating environmental conditions.

3.4 Adaptive interpretation and implications for grassland resilience

The interpretation of epigenetic variation in natural populations requires careful distinction between environmentally induced responses and genuinely adaptive modifications. Not all drought-induced epigenetic changes necessarily contribute to improved fitness. Therefore, adaptive significance should only be inferred when epigenetic features are consistently associated with drought tolerance across multiple accessions, persist following stress recovery, reappear during repeated stress cycles, correlate with environmental conditions at collection sites, or are experimentally validated through functional analyses (Springer and Schmitz, 2017; Fortes and Gallusci, 2017).

Several candidate regulatory modules are expected to play central roles in drought adaptation in S. viridis. These include ABA biosynthesis and signaling pathways, aquaporin-mediated water transport systems, late embryogenesis abundant (LEA) proteins, reactive oxygen species scavenging enzymes, and developmental regulators controlling flowering and reproductive success under water limitation. In addition, epigenetic regulators such as MET1, CMT2, CMT3, DRM2, AGO4, DDM1, and ROS1-like demethylases may directly influence the establishment and maintenance of adaptive chromatin states (Du et al., 2015; Zemach et al., 2013). Collectively, these findings highlight the potential of S. viridis as a powerful model for understanding how epigenetic mechanisms contribute to ecological resilience and drought adaptation in wild grass species.

4 Conservation, Epibreeding, and Future Directions

The conservation implications of drought epigenomics in wild grasses are substantial. If wild populations contain adaptive combinations of genotype and epigenotype shaped by local drought regimes, then conservation units should not be defined solely by neutral genetic structure. Instead, sampling strategies should deliberately capture climatic heterogeneity, especially along rainfall, aridity, and disturbance gradients. For seed banking and restoration, this means recording source moisture conditions, maternal environment, and regeneration procedures, because one practical risk is that ex situ propagation under benign conditions may erase or dilute ecologically relevant epigenetic states even when DNA sequence diversity is retained. In degraded drylands and encroached grasslands, restoration success may therefore depend not only on selecting locally adapted genotypes, but also on preserving environmentally calibrated regulatory states.

Epibreeding is the translational extension of this logic. In wild and semi-wild grasses, epibreeding should not be framed as replacing conventional breeding, but as adding a regulatory dimension to it. Three routes look most promising. The first is selection on stable natural epialleles that co-segregate with drought performance. The second is the deliberate use of stress priming, recurrent drought selection, or synthetic epigenetic populations to enrich favorable regulatory states before introgression or deployment. The third is epigenomic prediction, in which methylation and accessibility profiles are incorporated into models of drought performance alongside genotype and climate-of-origin data. This is especially attractive in complex, TE-rich grass genomes where many adaptive effects may be regulatory rather than coding.

Targeted epigenome editing offers a more direct future route. The RdDM literature now makes clear that locus-specific methylation can be induced in plants by engineered RNAs, hairpin constructs, or direct tethering of methylation machinery to specific loci, including CRISPR-based approaches. This is conceptually important for drought adaptation because it allows researchers to test whether methylation at a candidate promoter, enhancer, or TE actually changes phenotype without altering the underlying DNA sequence. For wild-grass research, that means moving from association to causality at candidate drought loci. In breeding terms, epigenome editing is

especially appealing where reversible or regulatory tuning is desired and where sequence modification is either unnecessary or undesirable.

Several challenges remain. Causality is still the central bottleneck. Many epigenetic marks mirror altered transcription instead of driving it. Tissue heterogeneity can blur signal, especially in leaves and meristems. Developmental stage strongly conditions the methylome and chromatin landscape, making time-resolved sampling essential. In grasses, repetitive DNA complicates mapping and interpretation. Population structure can also produce false adaptive signals if methylation is analyzed without genetics. Finally, transgenerational claims require particular caution because many stress-associated marks are reset, diluted, or contingent on repeated exposure. A rigorous field will therefore require pangenome-aware references, repeated-stress designs, recovery sampling, and explicit modeling of genotype, environment, and life history.

The most productive research agenda for 2026 and beyond is, in my view, clear. It should combine wild-population sampling across eco-climatic mosaics, long-read or pangenome-enabled methylome mapping, accessibility and histone profiling in stress time series, small-RNA analysis, fitness-linked common gardens, and targeted validation in tractable model grasses such as Setaria. That agenda is sufficiently mechanistic for molecular ecology, sufficiently ecological for conservation biology, and sufficiently translational for epibreeding.

Figure 1 showing how epigenetic information can inform seed collection, drought-resilient restoration, conservation prioritization, adaptive management, and regulatory breeding pipelines.

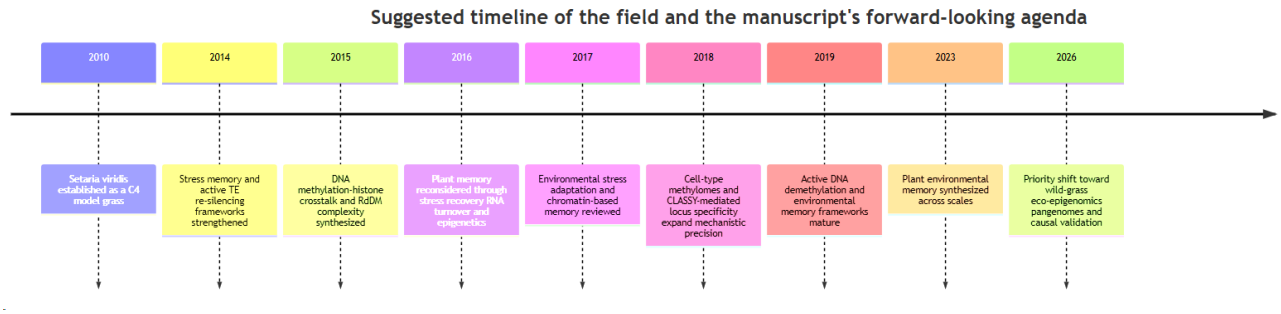

Figure 1 Conservation and epibreeding decision framework

5 Conclusions and Open Questions

Drought adaptation in wild grasses is not explained adequately by DNA sequence variation alone. The literature now supports a layered model in which drought cues are filtered through methylation pathways, histone-state transitions, accessibility changes, and small-RNA systems that collectively shape transcription, physiology, and potentially stress memory. This is especially relevant in grasses because of their ecological dominance in drylands and the regulatory importance of TE-rich genomes. The most secure mechanistic conclusions concern methylation pathway architecture, RdDM-mediated TE and stress regulation, and coupling between heterochromatin marks and gene control.

The main limitation of the current evidence base is taxonomic unevenness. Direct, multi-layer drought epigenomic datasets remain denser in model plants and crop systems than in truly wild grass populations, and integrated WGBS–ATAC–histone–RNA drought atlases are still rare for wild grasses. For that reason, Setaria viridis emerges as the most practical near-term organism for deriving causal, ecologically relevant insight that can later be transferred to less tractable taxa.

The most pressing open questions are these: Which epigenetic changes in wild grasses are adaptive rather than reactive? How stable are drought-associated states across development, dormancy, and seed regeneration? What fraction of field-relevant variation is controlled by cis-acting genetic differences versus environmentally induced epigenetic plasticity? How important are TE-adjacent regulatory changes in repeat-rich grass genomes? And can conservation and breeding programs deliberately preserve or induce beneficial regulatory states without

unacceptable instability or off-target effects? Those questions define the next phase of research at the boundary of molecular ecology, conservation, and climate adaptation.

Reference

Auge G., Hankofer V., Groth M., Antoniou-Kourounioti R., Ratikainen I., and Lampei C., 2023, Plant environmental memory: implications, mechanisms and opportunities for plant scientists and beyond, AoB Plants, 15(4): plad032.
https://doi.org/10.1093/aobpla/plad032

Bewick A.J., Niederhuth C.E., Ji L., Rohr N.A., Griffin P.T., Leebens-Mack J., and Schmitz R.J., 2017, The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants, Genome Biology, 18: 65.
https://doi.org/10.1186/s13059-017-1195-1

Blevins T., Pontvianne F., Cocklin R., Podicheti R., Chandrasekhara C., Yerneni S., Braun C., Lee B., Rusch D., Mockaitis K., Tang H.X., and Pikaard C.S., 2014, A two-step process for epigenetic inheritance in Arabidopsis, Molecular Cell, 54(1): 30-42.
https://doi.org/10.1016/j.molcel.2014.02.019

Brutnell T.P., Wang L., Swartwood K., Goldschmidt A., Jackson D., Zhu X.G., Kellogg E., and Van Eck J., 2010, Setaria viridis: a model for C4 photosynthesis, The Plant Cell, 22(8): 2537-2544.
https://doi.org/10.1105/tpc.110.075309

Cavrak V.V., Lettner N., Jamge S., Kosarewicz A., Bayer L.M., and Mittelsten Scheid O., 2014, How a retrotransposon exploits the plant's heat stress response for its activation, PLoS Genetics, 10: e1004115.
https://doi.org/10.1371/journal.pgen.1004115

Crisp P.A., Ganguly D., Eichten S.R., Borevitz J.O., and Pogson B.J., 2016, Reconsidering plant memory: intersections between stress recovery, RNA turnover, and epigenetics, Science Advances, 2: e1501340.
https://doi.org/10.1126/sciadv.1501340

Du J., Johnson L.M., Jacobsen S.E., and Patel D.J., 2015, DNA methylation pathways and their crosstalk with histone methylation, Nature Reviews Molecular Cell Biology, 16: 519-532.
https://doi.org/10.1038/nrm4043

Enke R.A., Dong Z., and Bender J., 2011, Small RNAs prevent transcription-coupled loss of histone H3 lysine 9 methylation in Arabidopsis thaliana, PLoS Genetics, 7: e1002350.
https://doi.org/10.1371/journal.pgen.1002350

Erdmann R.M., and Lafontaine Picard C., 2020, RNA-directed DNA methylation, PLoS Genetics, 16: e1009034.
https://doi.org/10.1371/journal.pgen.1009034

Fortes A.M., and Gallusci P., 2017, Plant stress responses and phenotypic plasticity in the epigenomics era: perspectives on the grapevine scenario, a model for perennial crop plants, Frontiers in Plant Science, 8: 82.
https://doi.org/10.3389/fpls.2017.00082

Gutzat R., and Mittelsten Scheid O., 2012, Epigenetic responses to stress: triple defense?, Current Opinion in Plant Biology, 15: 568-573.
https://doi.org/10.1016/j.pbi.2012.08.007

Hauser M.T., Aufsatz W., Jonak C., and Luschnig C., 2011, Transgenerational epigenetic inheritance in plants, Biochimica et Biophysica Acta Gene Regulatory Mechanisms, 1809: 459-468.
https://doi.org/10.1016/j.bbagrm.2011.03.007

Ito H., Gaubert H., Bucher E., Mirouze M., Vaillant I., and Paszkowski J., 2011, An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress, Nature, 472: 115-119.
https://doi.org/10.1038/nature09861

Jiang H., Barbier H., and Brutnell T., 2013, Methods for performing crosses in Setaria viridis, a new model system for the grasses, Journal of Visualized Experiments, 80: e50527.
https://doi.org/10.3791/50527-v

Lämke J., and Bäurle I., 2017, Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants, Genome Biology, 18: 124.
https://doi.org/10.1186/s13059-017-1263-6

Law J.A., and Jacobsen S.E., 2010, Establishing, maintaining and modifying DNA methylation patterns in plants and animals, Nature Reviews Genetics, 11: 204-220.
https://doi.org/10.1038/nrg2719

Matzke M.A., and Mosher R.A., 2014, RNA-directed DNA methylation: an epigenetic pathway of increasing complexity, Nature Reviews Genetics, 15: 394-408.
https://doi.org/10.1038/nrg3683

Parrilla-Doblas J.T., Roldán-Arjona T., Ariza R.R., and Córdoba-Cañero D., 2019, Active DNA demethylation in plants, International Journal of Molecular Sciences, 20: 4683.
https://doi.org/10.3390/ijms20194683

Popova O.V., Dinh H.Q., Aufsatz W., and Jonak C., 2013, The RdDM pathway is required for basal heat tolerance in Arabidopsis, Molecular Plant, 6: 396-410.
https://doi.org/10.1093/mp/sst023

Sebastian J., Wong M.K., Tang E., Dinneny J.R., and Brutnell T.P., 2014, Methods to promote germination of dormant Setaria viridis seeds, PLoS ONE, 9: e95109.
https://doi.org/10.1371/journal.pone.0095109

Sigman M.J., and Slotkin R.K., 2016, The first rule of plant transposable element silencing: location, location, location, The Plant Cell, 28: 304-313.
https://doi.org/10.1105/tpc.15.00869

Springer N.M., and Schmitz R.J., 2017, Exploiting induced and natural epigenetic variation for crop improvement, Nature Reviews Genetics, 18: 563-575.
https://doi.org/10.1038/nrg.2017.45

Tamiru M., Hardcastle T.J., and Lewsey M.G., 2018, Regulation of genome-wide DNA methylation by mobile small RNAs, New Phytologist, 217: 540-546.
https://doi.org/10.1111/nph.14874

Tricker P.J., Gibbings J.G., Rodríguez López C.M., Hadley P., and Wilkinson M.J., 2012, Low relative humidity triggers RNA-directed de novo DNA methylation and suppression of genes controlling stomatal development, Journal of Experimental Botany, 63: 3799-3813.
https://doi.org/10.1093/jxb/ers076

Wicker T., Gundlach H., Spannagl M., Uauy C., Borrill P., Ramírez-González R.H., De Oliveira R., International Wheat Genome Sequencing Consortium, Mayer K.F.X., Paux E., and Choulet F., 2018, Impact of transposable elements on genome structure and evolution in bread wheat, Genome Biology, 19(1): 103.
https://doi.org/10.1186/s13059-018-1479-0

Xu R., Wang Y.H., Zheng H., Lu W., Wu C.A., Huang J.G., Yan K., Yang G.D., and Zheng C.C., 2015, Salt-induced transcription factor MYB74 is regulated by the RNA-directed DNA methylation pathway in Arabidopsis, Journal of Experimental Botany, 66(19): 5997-6008.
https://doi.org/10.1093/jxb/erv312

Yang D.L., Zhang G.P., Wang L.L., Li J.W., Xu D.C., Di C.R., Tang K., Yang L., Zeng L., Miki D., Duan C.G., Zhang H.M., and Zhu J.K., 2018, Four putative SWI2/SNF2 chromatin remodelers have dual roles in regulating DNA methylation in Arabidopsis, Cell Discovery, 4(1): 55.
https://doi.org/10.1038/s41421-018-0056-8

Zemach A., Kim M.Y., Hsieh P.H., Coleman-Derr D., Eshed-Williams L., Thao K., Harmer S.L., and Zilberman D., 2013, The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin, Cell, 153(1): 193-205.
https://doi.org/10.1016/j.cell.2013.02.033

Zhao T., Zhan Z., and Jiang D., 2019, Histone modifications and their regulatory roles in plant development and environmental memory, Journal of Genetics and Genomics, 46: 467-476.
https://doi.org/10.1016/j.jgg.2019.09.005

Zhao Y., and Chen X., 2014, Non-coding RNAs and DNA methylation in plants, National Science Review, 1: 219-229.
https://doi.org/10.1093/nsr/nwu003

EcoEvo Publisher©

Reasons to publish in EcoEvo Publisher

An Online Publishing Platform

★Peer review quickly and professionally

☆Publish online immediately upon acceptance

★Deposit permanently and track easily

☆Access free and open around the world

★Disseminate multilingual available

Submit your manuscript at: **http://ecoevopublisher.com/**