

# Long-read sequencing in ecology and evolution: Understanding how complex genetic and epigenetic variants shape biodiversity

## 1 | INTRODUCTION

Ten years ago, the journal *Molecular Ecology* published a “road map” paper that reviewed past achievements in the discipline of molecular ecology, identified research challenges and charted a way forward (Andrew et al., 2013). That paper was motivated by a symposium organized during the First Joint Congress on Evolutionary Biology (Ottawa, July 6–10, 2012). In addition, it occurred on the heels of a major inflection point in molecular ecology and in life sciences more broadly: the development and uptake of “next”- or “second”-generation sequencing technologies, which deliver short DNA reads (typically shorter than 400 bp) at very high throughput (e.g., several billion reads per run; Goodwin et al., 2016). As such, Andrew et al. (2013) emphasized the promise of second-generation sequencing for diverse subdisciplines of molecular ecology such as phylogeography, landscape genomics, molecular adaptation and speciation. Representing more than just a technical advancement, second-generation sequencing was predicted to stimulate rapid conceptual breakthroughs in the field, especially in nonmodel species (Stapley et al., 2010; Tautz et al., 2010). As illustrated by any recent issue in the *Molecular Ecology* journal, these predictions were accurate.

While second-generation sequencing has enabled important discoveries at the forefront of molecular ecology, this technology does not come without limitations, the most prominent of which is short read length. Indeed, without additional validation, standard short reads cannot be used to traverse complex regions of the genome such as repetitive elements, duplications, inversions and other forms of structural change (Goodwin et al., 2016; Huddleston et al., 2014). Consequently, these regions have remained relatively unexplored. Ironically, however, they may also be particularly important for understanding ecological and evolutionary processes (Wellenreuther et al., 2019), given their high mutation rates (e.g., Hastings et al., 2009), and the fact that they can be extremely abundant—often surpassing single nucleotide polymorphisms by several fold, in terms of the total length of the genome that is affected (1000 Genomes Project Consortium et al., 2015; Mérot et al., 2023).

The technology to obtain reads spanning tens to thousands of kilobases was already available 10 years ago (Hayden, 2009; Munroe & Harris, 2010), and has been developed by several sequencing providers, the best-known of which are Pacific Biosciences (hereafter “PacBio”) and Oxford Nanopore Technologies (hereafter

“Nanopore”). Also referred to as third-generation sequencing, these approaches were initially used to complement short-read data (Goodwin et al., 2016; Laszlo et al., 2014; Munroe & Harris, 2010). Despite the potential utility of long reads, the adoption of this technology has been slow, primarily due to the high error rates reported for initial iterations of third-generation sequencing instruments (Glenn, 2011; Ip et al., 2015).

Recent years have brought about two important developments. First, the error rates of long reads have dropped considerably, in some cases below 1%, approaching rates characteristic of short reads (Goodwin et al., 2016). This is due to improvements in sequencing chemistry, base-calling algorithms and methods for post-sequencing error correction (Logsdon et al., 2020; Rang et al., 2018). Second, methods for assembling long stretches of DNA from short reads have also become available (e.g., McCoy et al., 2014; Selvaraj et al., 2013; Zheng et al., 2016). Collectively, these achievements are helping biologists tackle highly complex and dynamic regions of the genome, which were largely inaccessible until just a few years ago. Perhaps most clearly, the consequential role of long-read sequencing for getting the job done is illustrated by the recent publication of the first complete, telomere-to-telomere human genome (Nurk et al., 2022), more than two decades after the genome of our species was first made available (International Human Genome Sequencing Consortium, 2001).

This *Molecular Ecology* Special Issue highlights ways in which molecular ecologists are utilizing long-read information to explore the ecological and evolutionary roles of repetitive or otherwise complex loci. The 19 articles that comprise this issue, covering a range of plant, animal, bacteria and virus study systems, are grouped into six sections, which we summarize below. In doing so, our goal is to emphasize some of the key findings of each study. We also highlight, where possible, important challenges that will need to be overcome in the coming years, before long-read sequencing realizes its full potential. We conclude by summarizing the underlying thread of this Special Issue: that complex genetic and epigenetic variation, while traditionally more difficult to study, can make a substantial contribution to processes such as adaptation and speciation. We anticipate that, with continued improvement in long-read sequencing, this area of molecular ecology will only continue to grow, shaping our understanding of downstream biodiversity consequences of complex variants.

## 2 | PROFILING COMPLEX EPIGENETIC MODIFICATIONS

“Epigenetics” refers to heritable changes in the expression of the genome that are achieved by means other than direct modification in DNA sequence (Bossdorf et al., 2008). While a range of epigenetic mechanisms are known, including DNA methylation, histone modifications or small RNAs, research in ecology and evolution has focused largely on DNA methylation, because it is characterized by increased stability over generations (Verhoeven et al., 2016). Among the different types of methylated nucleotides, 5-methylcytosine (5mC) has received the most attention, as it is the dominant methylation pattern in eukaryotes (Goll & Bestor, 2005). Recent epigenomic studies have indicated that differential methylation can have wide-ranging ecological and evolutionary relevance. For example, broad methylation repatterning is known to follow hybridization and changes in the genomic background (Rapp & Wendel, 2005). As well, 5mC variants have been found to be associated with diverse environmental variables and with complex phenotypic or metabolic traits in a range of plant and animal species (Bossdorf et al., 2008; Hu & Barrett, 2017; Rapp & Wendel, 2005; Verhoeven et al., 2016). To obtain genome-wide profiles of 5mC in population samples, one may treat DNA with bisulphite prior to sequencing, in a step that converts unmethylated cytosines to uracil, rendering methylated vs. unmethylated cytosines identifiable in downstream sequence data (Verhoeven et al., 2016).

In one of the two opinion articles of this Special Issue, Nielsen et al. (2023) explore how long-read sequencing is revolutionizing epigenomic studies, using as an example bacteria and bacteriophages, which have a more diverse methylation repertoire than eukaryotes. The authors discuss several advantages that PacBio and Nanopore data offer for the detection of nucleotide modifications, including the fact that these technologies eliminate the need for bisulphite treatment and enable *de novo* detection of complex epigenetic base modifications. Aside from illustrating state-of-the-art approaches to data acquisition and analysis, Nielsen et al. (2023) also identify current limitations of epigenetic profiling as enabled by long-read sequencing, including the need to develop dedicated analytical tools that minimize noise from neighbouring nucleotides, and that implement reference libraries with the signature of diverse nucleotide modifications. As these advances are achieved, we will be much better positioned to understand the ecological and evolutionary relevance of diverse epigenetic modifications, including for the ongoing arms race between bacteria and bacteriophages (Nielsen et al., 2023).

## 3 | LINKING COMPLEX GENETIC LOCI TO GENOMIC NOVELTY

Repetitive regions of the genome such as transposable elements can be an important source of genomic novelty. Well-known routes to the reshuffling of chromosomal segments that involve

repetitive DNA include ectopic recombination and nonhomologous end-joining, which can lead to a variety of outcomes such as deletions, duplications, inversions or fusions (González & Petrov, 2012; Huang & Rieseberg, 2020). As illustrated by contributions included in this section, long-read sequencing can benefit investigations of how the repeat landscape may lead to changes in the karyotype or in patterns of synteny. As an example of changes in karyotype, Burley et al. (2023) use Nanopore long reads to characterize a large (134-Mbp) neo-sex chromosome in the blue-faced honeyeater. Results demonstrated that this chromosome originated via a fusion between an autosome and the ancestral Z chromosome, with important consequences for the genomic landscape of diversity and differentiation. Remarkably, the same chromosomal regions appear to have fused convergently in other songbird lineages, potentially facilitated by repeats that are shared between the two chromosomes (Burley et al., 2023). As an example of changes in synteny, Ferguson et al. (2023) use Nanopore long reads to sequence, assemble and compare the genomes of three *Eucalyptus* species. Results demonstrated that transposon-rich regions of the genome can lead to synteny loss via small-scale rearrangements. Their study thus challenges the generally accepted view that *Eucalyptus* species maintain a largely syntenic genome. Moreover, results showed that a sizeable fraction of rearrangements contained genes, and therefore have the potential to drive adaptation in this species-rich and widely distributed genus.

Rather than representing an obstacle to be overcome during genome assembly, repetitive DNA may also be the focus of study. Peona et al. (2023), for example, investigated the evolution of satellite repetitive DNA in 24 species of birds. Using linked short reads and PacBio long reads, the authors catalogued repeats with monomer sizes ranging from 20 bp to 4 kb that are highly dynamic. Remarkably, patterns of satellite DNA abundance did not align with predictions of current models for satellite DNA evolution. Specifically, satellite DNA profiles were found to be more similar among deeply diverged species than among recently diverged species. This result therefore highlights a promising area for future study. In addition, Wierzbicki et al. (2023) investigated piRNA (PIWI-interacting RNA) clusters in *Drosophila*. These genomic clusters are known to be rich in repetitive elements and have a crucial role in the genomic defence against transposable elements. The authors resolved 20 such clusters from four *Drosophila* species, using contiguous genome assemblies made with PacBio and Nanopore data. Aside from developing a framework for quantitative investigations of the dynamics of piRNA clusters, which includes establishing synteny between these highly dynamic loci, Wierzbicki et al. (2023) show that piRNA clusters evolve rapidly, mainly due to the insertion of recent transposable elements, and the deletion of old ones. Remaining challenges include expanding the taxonomic breadth of studies of piRNA cluster evolution, as well as extending analyses to a larger fraction of the total piRNA complement of each genome. As Wierzbicki et al. (2023) point out, both challenges stand to be overcome with the increased implementation of long-read sequencing in molecular ecology.

## 4 | CONSERVATION GENOMICS

Long-read sequencing is broadening the toolset available for the management of small populations, by enabling the reconstruction of gap-free, highly contiguous genomes at a fast pace (Kardos et al., 2021). This allows us to re-evaluate previous conclusions that were drawn for at-risk populations based on genetic data, such as population origin, levels of inbreeding or genetic structure (Kardos et al., 2021). Moreover, genomic data sets also allow new information to be gained regarding the long-term demographic and evolutionary histories of populations, or the contribution of structural variants to population fitness (Kardos et al., 2021; Wold et al., 2021).

This Special Issue includes examples of long-read genome-scale analyses in vulnerable or threatened species. For instance, Li, Yang, et al. (2023) rely on PacBio data and Hi-C technology to assemble the genome of the takin, a large bovid herbivore currently listed as vulnerable by the International Union for Conservation of Nature (IUCN; Li, Yang, et al., 2023). This high-quality chromosome-level assembly was used, along with resequencing data, to demonstrate important declines in effective population size during the past million years, and to uncover evidence of runs of homozygosity caused by recent inbreeding. In another example, Yan et al. (2023) investigate intraspecific divergence in a hot-spring snake that is endemic to the Qinghai-Tibet Plateau and is listed as near threatened by IUCN. The authors use short-read data to infer intraspecific divergence, reconstruct demographic history and find genes under selection during local adaptation. By combining these data with PacBio long reads, the authors are also able to document the abundance of structural variants and assess their contribution to differentiation among major lineages in this system.

## 5 | ADAPTATION GENOMICS

Being able to thoroughly catalogue genetic diversity is critically important for answering some of the most fundamental questions in evolutionary biology, such as how wild populations are likely to respond when confronted with challenging or novel environments (e.g., Yeaman et al., 2016), or whether and why adaptive evolution repeatedly makes use of the same genetic modules (e.g., Jones et al., 2012). As demonstrated by a number of contributions in this Special Issue, long-read sequencing is helping us answer these questions.

Xie et al. (2023), for instance, study how mangroves cope with a unique environment: the interface of land and sea. The authors rely on a combination of short reads and PacBio long reads to obtain chromosome-level assemblies for two mangrove species, and for one closely related inland species. In contrast to previous studies in other mangroves, which found that whole genome duplications preceded the colonization of novel habitats, Xie et al. (2023) do not detect evidence of recent polyploidization. Rather, they attribute the large genome sizes of these species to repeat sequence expansion. Additionally, results emphasized lack of parallelism in gene family evolution, consistent with the use of different genetic modules

during adaptation to the intertidal environment in these species. Evidence for repeated use of the same functional genes was found, however, in the study of Li, Wang, et al. (2023). The authors relied on a new chromosome-level assembly for a tropical poplar, obtained with Nanopore long reads. Comparisons with five other poplar species provided evidence of convergent evolution during adaptation to tropical environments. Hotaling et al. (2023) provide another exciting example of how long reads are fast-tracking the study of adaptation. The authors relied on PacBio long reads to obtain a genome assembly for the Antarctic eelpout, the first representative of the family Zoarcidae of ray-finned fish to be genome-sequenced. This highly contiguous assembly in turn allowed the authors to focus on regions of the genome such as the haemoglobin and antifreeze gene clusters which, while representing strong candidates for cold water adaptation, are arranged in highly duplicated tandem arrays (Hotaling et al., 2023). Results were consistent with convergent as well as species-specific mechanisms of adaptation to the extremely cold waters of the Southern Ocean.

A series of other studies in this section illustrate the utility of long reads for understanding the genetic architecture of functionally important traits. Nacif et al. (2023), for example, conduct a comprehensive investigation of the sex-determining region in Midas cichlid fish. Using a combination of forward-genetics, PacBio sequencing and Bionano optical mapping, the authors narrow down sex determination in this system to an ~100-kb region of the Y chromosome that is rich in transposable elements. This region harbours a few partial genes, but also one complete coding gene: a duplicate of the anti-Mullerian receptor 2 gene (*amhr2Y*). Because *amhr2Y* has been shown to act as a molecular sex-determining locus in other teleost fishes (Nacif et al., 2023), it represents a strong candidate for future functional validation, and probably an additional example of molecular parallelism. At the other extreme in terms of the scale of duplication, Zhu et al. (2023) focus on a biennial alpine plant that sustained two recent rounds of whole genome duplication. In this system, the authors detail a multi-omics investigation of dimorphic cleistogamy. Known to have evolved repeatedly across plants, dimorphic cleistogamy is manifested by the production of both open (available for cross-pollination) and closed (self-fertilizing) flowers, and as such is thought to be important for reproductive assurance in challenging environments (Zhu et al., 2023). An assembly made using short reads and Nanopore data revealed a genome that consists of over 70% repetitive sequences. By integrating additional experiments that probed changes in gene expression and in metabolites, the authors were able to show that a large number of genes and metabolites differentiate the two types of flowers. This is consistent with a complex genetic architecture for this trait, which can at least partially be attributed to past whole genome duplication events (Zhu et al., 2023). Finally, Cohen et al. (2023) investigate the genetics of pesticide resistance in Colorado potato beetle. The authors used PacBio sequencing and a trio-binning approach to obtain three new haploid assemblies with considerably improved contiguity, as compared to the existing reference genome for this pest species. A pangome obtained using all assemblies as well as population-scale

resequencing data are then used to investigate the role of structural variants in rapid adaptation to pesticide exposure. Results revealed that structural variants are abundant, accounting for ~30% of the genome, while also highlighting cases in which structural variants may have been adaptive.

Such studies demonstrate the relevance of long-read sequencing for understanding the process of adaptation. At the same time, they underline a need for developing analytical approaches that are designed for structural variants, and that additionally exploit other layers of information made available by third-generation sequencing. For example, in the second opinion article of this Special Issue, Shipilina et al. (2023) consider and illustrate, based on simulated and empirical data, the utility of haplotype information that can be obtained using long-read technology, including for analyses of selective sweeps. Specifically, the authors discuss methods based on ancestral recombination graph reconstruction, which, in addition to mutation, take into account ancestry and recombination. This information, while currently computationally challenging to obtain for large numbers of samples, could vastly improve resolution as compared to data sets based only on single nucleotide polymorphisms, including by identifying multiple selective sweeps that occur in the same genomic region (Shipilina et al., 2023).

## 6 | GENOMICS OF SPECIATION AND HYBRIDIZATION

As adaptation proceeds and populations diverge, reproductive isolation may gradually develop. Genomic analyses of species pairs, and in particular of those pairs that have recently diverged, represent a promising approach for dissecting the genetic architecture of speciation and for identifying barrier loci (Ravinet et al., 2017). Several papers in this Special Issue focus on the speciation continuum and investigate the contribution of structural variants and recombination suppression to species differentiation. Mérot et al. (2023), for example, undertake a detailed genomic investigation of recent speciation using a pair of whitefish species that diverged in allopatry starting around 60,000 years ago, and then came back into contact roughly 12,000 years ago (Mérot et al., 2023). The authors combined short-read resequencing with Nanopore long reads to obtain the first genome assemblies for both Dwarf and Normal whitefish species, and to genotype single nucleotide polymorphisms and structural variants. Whitefish genomes were found to be repeat-rich, with over 60% of sequence corresponding to interspersed repeats. Moreover, results indicated that a large proportion of the structural variants that differentiate the two species were enriched for several classes of transposable elements. This is consistent with a role of bursts in repetitive elements in generating early genome-wide differentiation between species, and even reproductive isolation (Mérot et al., 2023).

In another investigation of incipient speciation, Wersebe et al. (2023) focus on the freshwater crustacean *Daphnia*. The authors present the first genome-wide scan of differentiation for the

*pulex–pulicaria* pair of species, which separated roughly 150,000 years ago (Wersebe et al., 2023). A reference genome made for *D. pulicaria* using PacBio data, complemented with short reads for both species and their hybrids, allowed the authors to reconstruct the genomic landscape of differentiation. Contrary to expectations, results indicated that genomic windows of high differentiation between these species are restricted to genic regions of high recombination. Finally, Zhang et al. (2023) present an in-depth investigation of the contribution of structural variation to reproductive isolation, in one of the few studies so far that has implemented population-scale long-read sequencing. The authors focus on a natural hybrid zone established between two species of *Lycaeides* butterflies that separated over 2.4 million years ago, and that came into secondary contact roughly 14,000 years ago (Zhang et al., 2023). Structural variants were genotyped for parental and hybrid individuals using Nanopore data, and then validated using short reads. Genomic cline analyses revealed over 562 structural variants with a signature of selection in the hybrid zone. Among different structural variants, deletions were found to exhibit the largest departures from neutral expectations, pointing to a large contribution of these variants, along with gene-rich inversions, to hybrid fitness and reproductive isolation (Zhang et al., 2023).

## 7 | GENOMICS OF SPECIES INTERACTIONS

In addition to facilitating in-depth analysis of epigenomes, genomes, populations and species, long-read sequencing can also be harnessed to study species interactions. As illustrated by the two papers in this section, this information can be broadly relevant in contexts that range from pathogen control to understanding how biological communities are assembled. In the first paper, van Steenbrugge et al. (2023) study the evolution of virulence in potato cyst nematodes, which are among the most destructive pathogens of potato worldwide. The authors rely on Nanopore data to assemble a new and highly contiguous reference genome for potato cyst nematodes as well as for an outgroup. These genomes are in turn used to investigate six families of effectors, which are proteins secreted by the pathogen that can manipulate plant physiology and have a key role in virulence. Aside from illuminating patterns of evolutionary diversification for effector genes, results are also predicted to facilitate the management of potato cyst nematodes. Specifically, the findings of van Steenbrugge et al. (2023) should enable molecular investigations of pathogen populations, and subsequent matching of potato host resistance genes with pathogen virulence genotypes. In the second paper in this section, Handy et al. (2023) rely on PacBio data to investigate the composition of gut bacterial communities for two carpenter bee species that are incipiently social. In this case, the ability to obtain full-length 16S amplicons via long-read sequencing allowed the authors to classify bacterial species with significantly improved resolution, and to reveal in this way species interactions that would have otherwise remained cryptic. Results revealed both shared and distinct elements of the microbiome between the two

bee species. Moreover, results indicated that different components of the microbiome might be structured by different processes, including geographical isolation and patterns of microbial transmission, highlighting a promising area for future investigation.

## 8 | CONCLUSIONS

As illustrated by the collection of articles in this Special Issue, long-read sequencing is providing molecular ecologists with the information needed to tackle some of the most challenging basic and applied topics in our discipline, with important discoveries being made across groups of organisms and levels of biological organization. These studies collectively emphasize the underlying thread of this Special Issue: epigenetic variants, structural variants, repetitive elements and other regions in the genome that may have been hard to assemble and genotype using short reads can now be properly traversed and can play a critical role during adaptation and species diversification. In this context, we emphasize the need for expanding long-read sequencing at scales that exceed single individuals. While obtaining highly contiguous reference genomes is an essential first step, we stand to gain much more from replicate genome assemblies, pangenomes and population-scale long-read sequencing, as illustrated by articles in this Special Issue. At the same time, there is a critical need for analytical improvements. These include developing methods that explicitly consider structural variation, as well as improving the computational efficiency of existing methods, such that long-range information provided by third-generation sequencing can be efficiently harnessed for large numbers of samples.

Ten years ago, in their road map paper, Andrew et al. (2013) emphasized how next-generation sequencing is improving our observational abilities, illuminating new areas of study. We are now in a similar position, on the cusp of accelerated progress facilitated by recent advances in long-read sequencing technology. With continued broadening of the molecular and analytical toolkits available to molecular ecologists, we are increasingly able to push the conceptual limits of our discipline, and to answer ever more challenging questions of basic and applied relevance about the biodiversity that sustains us and surrounds us.

### ACKNOWLEDGEMENTS

We would like to acknowledge all the authors who contributed articles to this Special Issue, the reviewers who evaluated the manuscripts, as well as editors at the *Molecular Ecology* journal including Emily Warschewsky and Ben Sibbett for their help throughout.

### DATA AVAILABILITY STATEMENT

Not applicable.

Dan G. Bock<sup>1</sup>   
 Jianquan Liu<sup>2,3</sup>   
 Polina Novikova<sup>4</sup>   
 Loren H. Rieseberg<sup>1</sup> 

<sup>1</sup>Department of Botany, Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada

<sup>2</sup>State Key Laboratory of Grassland and Agro-ecosystems, Institute of Innovation Ecology, School of Life Science and the Supercomputing Center, Lanzhou University, Lanzhou, China

<sup>3</sup>Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China

<sup>4</sup>Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany

### Correspondence

Dan G. Bock, Department of Botany, Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada.

Email: [dan.g.bock@gmail.com](mailto:dan.g.bock@gmail.com)

Handling Editor: Benjamin Sibbett

### ORCID

Dan G. Bock  <https://orcid.org/0000-0001-7788-1705>

Jianquan Liu  <https://orcid.org/0000-0002-4237-7418>

Polina Novikova  <https://orcid.org/0000-0002-4552-4575>

Loren H. Rieseberg  <https://orcid.org/0000-0002-2712-2417>

### REFERENCES

- 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, 526, 68–74.
- Andrew, R. L., Bernatchez, L., Bonin, A., Buerkle, C. A., Carstens, B. C., Emerson, B. C., Garant, D., Giraud, T., Kane, N. C., Rogers, S. M., Slate, J., Smith, H., Sork, V. L., Stone, G. N., Vines, T. H., Waits, L., Widmer, A., & Rieseberg, L. H. (2013). A road map for molecular ecology. *Molecular Ecology*, 22, 2605–2626. <https://doi.org/10.1111/mec.12319>
- Bossdorf, O., Richards, C. L., & Pigliucci, M. (2008). Epigenetics for ecologists. *Ecology Letters*, 11, 106–115.
- Burley, J. T., Orzechowski, S., Sin, S., & Edwards, S. V. (2023). Whole-genome phylogeography of the blue-faced honeyeater (*Entomyzon cyanotis*) and discovery and characterization of a neo-Z chromosome. *Molecular Ecology*, 32, 1248–1270. <https://doi.org/10.1111/mec.16604>
- Cohen, Z. P., Schoville, S. D., & Hawthorne, D. J. (2023). The role of structural variants in pest adaptation and genome evolution of the Colorado potato beetle, *Leptinotarsa decemlineata* (Say). *Molecular Ecology*, 32, 1425–1440. <https://doi.org/10.1111/mec.16838>
- Ferguson, S., Jones, A., Murray, K., Schwesinger, B., & Borevitz, J. O. (2023). Interspecies genome divergence is predominantly due to frequent small scale rearrangements in eucalyptus. *Molecular Ecology*, 32, 1271–1287. <https://doi.org/10.1111/mec.16608>
- Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, 11, 759–769. <https://doi.org/10.1111/j.1755-0998.2011.03024.x>
- Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry*, 74, 481–514. <https://doi.org/10.1146/annurev.biochem.74.010904.153721>
- González, J., & Petrov, D. A. (2012). Evolution of genome content: Population dynamics of transposable elements in flies and humans.

- In M. Anisimova (Ed.), *Evolutionary genomics. Methods in molecular biology (methods and protocols)* (Vol. 855, pp. 361–383). Humana Press.
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17, 333–351. <https://doi.org/10.1038/nrg.2016.49>
- Handy, M. Y., Sbardellati, D. L., Yu, M., Saleh, N. W., Ostwald, M. M., & Vannette, R. L. (2023). Incipiently social carpenter bees (*Xylocopa*) host distinctive gut bacterial communities and display geographical structure as revealed by full-length PacBio 16S rRNA sequencing. *Molecular Ecology*, 32, 1530–1543. <https://doi.org/10.1111/mec.16736>
- Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. *Nature Reviews Genetics*, 10, 551–564.
- Hayden, E. (2009). Genome sequencing: The third generation. *Nature*, 457, 768–769.
- Hotaling, S., Desvignes, T., Sproul, J. S., Lins, L. S. F., & Kelley, J. L. (2023). Pathways to polar adaptation in fishes revealed by long-read sequencing. *Molecular Ecology*, 32, 1381–1397. <https://doi.org/10.1111/mec.16501>
- Hu, J., & Barrett, R. D. H. (2017). Epigenetics in natural animal populations. *Journal of Evolutionary Biology*, 30, 1612–1632. <https://doi.org/10.1111/jeb.13130>
- Huang, K., & Rieseberg, L. H. (2020). Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Frontiers in Plant Science*, 11, 296. <https://doi.org/10.3389/fpls.2020.00296>
- Huddleston, J., Ranade, S., Malig, M., Antonacci, F., Chaisson, M., Hon, L., Sudmant, P. H., Graves, T. A., Alkan, C., Dennis, M. Y., Wilson, R. K., Turner, S. W., Korf, J., & Eichler, E. E. (2014). Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Research*, 24, 688–696.
- International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409, 860–921. <https://doi.org/10.1038/35057062>
- Ip, C. L. C., Loose, M., Tyson, J., de Cesare, M., Brown, B. L., Jain, M., Leggett, R. M., Eccles, D. A., Zalunin, V., Urban, J. M., Piazza, P., Bowden, R. J., Paten, B., Mwaigwisya, S., Batty, E. M., Simpson, J. T., Snutch, T. P., Birney, E., Buck, D., ... MinION Analysis and Reference Consortium. (2015). MinION Analysis and Reference Consortium: Phase 1 data release and analysis [version 1; Referees, approved]. *F1000Research*, 4, 1075.
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., ... Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484, 55–61.
- Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., Tallmon, D. A., & Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences of the United States of America*, 118, e2104642118. <https://doi.org/10.1073/pnas.2104642118>
- Laszlo, A. H., Derrington, I. M., Ross, B. C., Brinkerhoff, H., Adey, A., Nova, I. C., Craig, J. M., Langford, K. W., Samson, J. M., Daza, R., Doering, K., Shendure, J., & Gundlach, J. H. (2014). Decoding long nanopore sequencing reads of natural DNA. *Nature Biotechnology*, 32, 829–833.
- Li, A., Yang, Q., Li, R., Dai, X., Cai, K., Lei, Y., Jia, K., Jiang, Y., & Zan, L. (2023). Chromosome-level genome assembly for takin (*Budorcas taxicolor*) provides insights into its taxonomic status and genetic diversity. *Molecular Ecology*, 32, 1323–1334. <https://doi.org/10.1111/mec.16483>
- Li, Y., Wang, D., Wang, W., Yang, W., Gao, J., Zhang, W., Shan, L., Kang, M., Chen, Y., & Ma, T. (2023). A chromosome-level *Populus qiongdaoensis* genome assembly provides insights into tropical adaptation and a cryptic turnover of sex determination. *Molecular Ecology*, 32, 1366–1380. <https://doi.org/10.1111/mec.16566>
- Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nature Reviews Genetics*, 21, 597–614. <https://doi.org/10.1038/s41576-020-0236-x>
- McCoy, R. C., Taylor, R. W., Blauwkamp, T. A., Kelley, J. L., Kertesz, M., Pushkarev, D., Petrov, D. A., & Fiston-Lavier, A.-S. (2014). Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. *PLoS One*, 9, e106689.
- Mérot, C., Stenlökk, K. S. R., Venney, C., Laporte, M., Moser, M., Normandeau, E., Árnýasi, M., Kent, M., Rougeux, C., Flynn, J. M., Lien, S., & Bernatchez, L. (2023). Genome assembly, structural variants, and genetic differentiation between lake whitefish young species pairs (*Coregonus* sp.) with long and short reads. *Molecular Ecology*, 32, 1458–1477. <https://doi.org/10.1111/mec.16468>
- Munroe, D., & Harris, T. (2010). Third-generation sequencing fireworks at Marco Island. *Nature Biotechnology*, 28, 426–428. <https://doi.org/10.1038/nbt0510-426>
- Nacif, C. L., Kratochwil, C. F., Kautt, A. F., Nater, A., Machado-Schiaffino, G., Meyer, A., & Henning, F. (2023). Molecular parallelism in the evolution of a master sex-determining role for the anti-Mullerian hormone receptor 2 gene (*amhr2*) in Midas cichlids. *Molecular Ecology*, 32, 1398–1410. <https://doi.org/10.1111/mec.16466>
- Nielsen, T. K., Forero-Junco, L. M., Kot, W., Moineau, S., Hansen, L. H., & Riber, L. (2023). Detection of nucleotide modifications in bacteria and bacteriophages: Strengths and limitations of current technologies and software. *Molecular Ecology*, 32, 1236–1247. <https://doi.org/10.1111/mec.16679>
- Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bizikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonzo, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., ... Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, 376, 44–53. <https://doi.org/10.1126/science.abj6987>
- Peona, V., Kutschera, V. E., Blom, M. P. K., Irestedt, M., & Suh, A. (2023). Satellite DNA evolution in Corvoidea inferred from short and long reads. *Molecular Ecology*, 32, 1288–1305. <https://doi.org/10.1111/mec.16484>
- Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, 19, 90.
- Rapp, R. A., & Wendel, J. F. (2005). Epigenetics and plant evolution. *New Phytologist*, 168, 81–91.
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30, 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Selvaraj, S., Dixon, J. R., Bansal, V., & Ren, B. (2013). Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nature Biotechnology*, 31, 1111–1118. <https://doi.org/10.1038/nbt.2728>
- Shipilina, D., Pal, A., Stankowski, S., Chan, Y. C., & Barton, N. H. (2023). On the origin and structure of haplotype blocks. *Molecular Ecology*, 32, 1441–1457. <https://doi.org/10.1111/mec.16793>
- Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A. D., Beckerman, A. P., & Slate, J. (2010). Adaptation genomics: The next generation. *Trends in Ecology and Evolution*, 25, 705–712. <https://doi.org/10.1016/j.tree.2010.09.002>
- Tautz, D., Ellegren, H., & Weigel, D. (2010). Next generation molecular ecology. *Molecular Ecology*, 19, 1–3. <https://doi.org/10.1111/j.1365-294X.2009.04489.x>
- van Steenbrugge, J. J. M., van den Elsen, S., Holterman, M., Lozano-Torres, J. L., Putker, V., Thorpe, P., Goverse, A., Sterken, M. G.,

- Smant, G., & Helder, J. (2023). Comparative genomics among cyst nematodes reveals distinct evolutionary histories among effector families and an irregular distribution of effector-associated promoter motifs. *Molecular Ecology*, 32, 1515–1529. <https://doi.org/10.1111/mec.16505>
- Verhoeven, K. J. F., vonHoldt, B. M., & Sork, V. L. (2016). Epigenetics in ecology and evolution: What we know and what we need to know. *Molecular Ecology*, 25, 1631–1638. <https://doi.org/10.1111/mec.13617>
- Wellenreuther, M., Mérot, C., Berdan, E., & Bernatchez, L. (2019). Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Molecular Ecology*, 28, 1203–1209. <https://doi.org/10.1111/mec.15066>
- Wersebe, M. J., Sherman, R. E., Jeyasingh, P. D., & Weider, L. J. (2023). The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex. *Molecular Ecology*, 32, 1478–1496. <https://doi.org/10.1111/mec.16383>
- Wierzbicki, F., Kofler, R., & Signor, S. (2023). Evolutionary dynamics of piRNA clusters in *Drosophila*. *Molecular Ecology*, 32, 1306–1322. <https://doi.org/10.1111/mec.16311>
- Wold, J., Koepfli, K.-P., Galla, S. J., Eccles, D., Hogg, C. J., Le Lec, M. F., Guhlin, J., Santure, A. W., & Steeves, T. E. (2021). Expanding the conservation genomics toolbox: Incorporating structural variants to enhance genomic studies for species of conservation concern. *Molecular Ecology*, 30, 5949–5965. <https://doi.org/10.1111/mec.16141>
- Xie, W., Guo, Z., Wang, J., He, Z., Li, Y., Feng, X., Zhong, C., & Shi, S. (2023). Evolution of woody plants to the land-sea interface – The atypical genomic features of mangroves with atypical phenotypic adaptation. *Molecular Ecology*, 32, 1351–1365. <https://doi.org/10.1111/mec.16587>
- Yan, C., Song, M.-H., Jiang, D., Ren, J.-L., Lv, Y., Chang, J., Huang, S., Zaher, H., & Li, J.-T. (2023). Genomic evidence reveals intraspecific divergence of the hot-spring snake (*Thermophis baileyi*), an endangered reptile endemic to the Qinghai-Tibet plateau. *Molecular Ecology*, 32, 1335–1350. <https://doi.org/10.1111/mec.16687>
- Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., Nurkowski, K. A., Smets, P., Wang, T., Gray, L. K., Liepe, K. J., Hamann, A., Holliday, J. A., Whitlock, M. C., Rieseberg, L. H., & Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science*, 353, 1431–1433. <https://doi.org/10.1126/science.aaf7812>
- Zhang, L., Chaturvedi, S., Nice, C. C., Lucas, L. K., & Gompert, Z. (2023). Population genomic evidence of selection on structural variants in a natural hybrid zone. *Molecular Ecology*, 32, 1497–1514. <https://doi.org/10.1111/mec.16469>
- Zheng, G. X. Y., Lau, B. T., Schnall-Levin, M., Jarosz, M., Bell, J. M., Hindson, C. M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D. A., Merrill, L., Terry, J. M., Mudivarti, P. A., Wyatt, P. W., Bharadwaj, R., Makarewicz, A. J., Li, Y., Belgrader, P., Price, A. D., Lowe, A. J., Marks, P., ... Ji, H. P. (2016). Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nature Biotechnology*, 34, 303–311. <https://doi.org/10.1038/nbt.3432>
- Zhu, M., Wang, Z., Yang, Y., Wang, Z., Mu, W., & Liu, J. (2023). Multi-omics reveal differentiation and maintenance of dimorphic flowers in an alpine plant on the Qinghai-Tibet plateau. *Molecular Ecology*, 32, 1411–1424. <https://doi.org/10.1111/mec.16449>