

# Neutral Theory Is the Foundation of Conservation Genetics

Anne D. Yoder,<sup>\*1</sup> Jelmer W. Poelstra,<sup>1</sup> George P. Tiley,<sup>1</sup> and Rachel C. Williams<sup>1</sup>

<sup>1</sup>Department of Biology, Duke University, Durham, NC

\*Corresponding author: E-mail: anne.yoder@duke.edu.

Associate editor: Sudhir Kumar

## Abstract

Kimura's neutral theory of molecular evolution has been essential to virtually every advance in evolutionary genetics, and by extension, is foundational to the field of conservation genetics. Conservation genetics utilizes the key concepts of neutral theory to identify species and populations at risk of losing evolutionary potential by detecting patterns of inbreeding depression and low effective population size. In turn, this information can inform the management of organisms and their habitat providing hope for the long-term preservation of both. We expand upon Avise's "inventorial" and "functional" categories of conservation genetics by proposing a third category that is linked to the coalescent and that we refer to as "process-driven." It is here that connections between Kimura's theory and conservation genetics are strongest. Process-driven conservation genetics can be especially applied to large genomic data sets to identify patterns of historical risk, such as population bottlenecks, and accordingly, yield informed intuitions for future outcomes. By examining inventorial, functional, and process-driven conservation genetics in sequence, we assess the progression from theory, to data collection and analysis, and ultimately, to the production of hypotheses that can inform conservation policies.

**Key words:** demography, cryptic species, coalescent methods, biodiversity, effective population size, genomics.

## Introduction

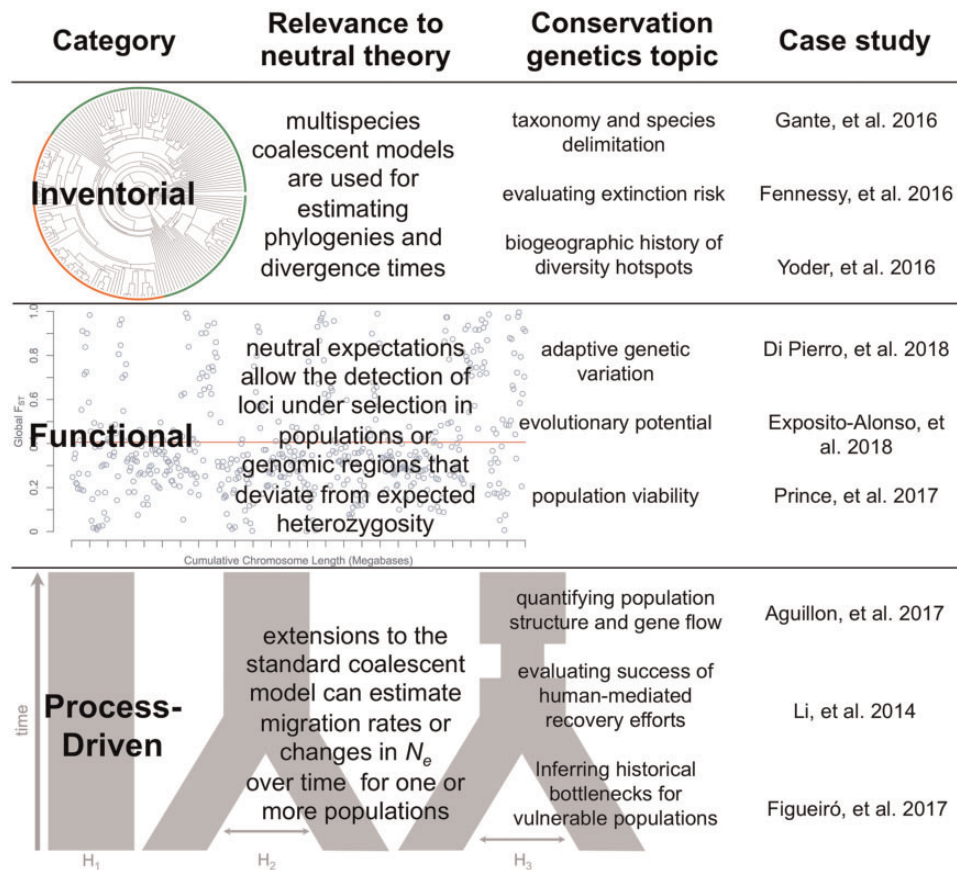
Kimura (1968) could not possibly have foreseen the profound impacts that his neutral theory of molecular evolution would have on virtually every field of evolutionary genetics. Certainly, he could not have imagined that, as we will argue here, the field of conservation genetics would be powerless without the theoretical underpinnings of neutral theory. In fact, Kimura's two principle arguments—1) that the majority of change at the genetic level is essentially invisible to natural selection and 2) that rates of genetic drift through time largely reflect the interaction of effective population size ( $N_e$ ) and generation time—have been absolutely foundational to every aspect of the conservation genetic enterprise. More specifically, one of the profound implications of these ideas was that for "mutant genes," the probability of fixation is proportional to its initial frequency, and thus, new alleles may be produced at the same rate per individual as they are substituted in the population (Kimura 1968). This insight in particular yields the conditions under which coalescent theory applies, consequently making the study of historical demography feasible. We will argue that this logical thread—from neutral theory, to the coalescent, to historical demography—has been a conceptual breakthrough that yields tremendous power for conservation genetics.

To draw this connection between neutral theory and conservation genetics is not novel (Lynch and O'Hely 2001), though we believe that the time is ripe to explicitly touch on the essentials of Kimura's theory and the myriad ways in which it informs every aspect of both operational and aspirational conservation genetics. Further, we draw attention to

the new power conferred by the genomics revolution of the past decade. Though a number of authors have examined the promises and pitfalls of genomic data applied to conservation practice (Primmer 2009; Ouborg et al. 2010; Steiner et al. 2013; Russello et al. 2015; Shafer et al. 2015; Benestan et al. 2016), we find Avise (2010) to be especially useful in the partitioning of conservation genetics into two categories: "inventorial" and "functional." Inventorial conservation genetics seeks to characterize and measure levels of species diversity, whereas functional conservation genetics seeks to measure the relative fitness of individuals and populations. We believe that the time has come to consider a third category, one that we regard as especially linked to neutral theory and the coalescent, which we refer to as "process-driven" (fig. 1). By examining the three categories in sequence, there is a framework for assessing the progression from theory to data collection and analysis, and ultimately, to the production of hypotheses that can inform conservation policies.

## The Goals of Conservation Genetics

We have now had 50 years to integrate the theoretical advances of Kimura's ideas into our conservation practices. Genetic variation is the raw material of evolutionary potential, and thus, the most obvious goal for conservation genetics is to characterize and propose measures for preserving this variation within populations and species. Therefore, maximizing evolutionary potential may be considered the unifying principle for all aspects of conservation genetic theory and practice. Conservation genetics applies molecular evolution to detect symptoms of inbreeding depression as measured by



**FIG. 1.** Select case studies that illustrate the relationship between Kimura's neutral theory of molecular evolution and conservation goals for inventorial, functional, and process-driven conservation genetics.

low levels of heterozygosity, high levels of identity by descent, and small  $N_e$ , as well as by estimating the rates and directionality of gene flow among populations. Earth now faces an extinction crisis unparalleled for 65 My (Ceballos et al. 2015). Increasingly known as the Sixth Extinction, the causes are entirely human-mediated, and thus, the solutions must also be. Conservation geneticists cannot directly enforce action to rescue populations, species, and ecosystems at risk, but we can provide the evidence needed to direct legislation for political leaders and organizations who are in the position to develop and implement policy.

### “Inventorial” Conservation Genetics

Up to the present, many conservation genetic studies have been inventorial in nature, and for good reason. The first goal of conservation biology is to assemble the data and conduct the analyses necessary to characterize the composition, complexity, and general evolutionary “health” of a given population or habitat. Inventorial conservation genetics encompasses quantitative insight into sequence variation at the individual, population, and species level. This is particularly important given that “current extinction risk is not random” with phylogenetic identity often being a strong predictor of threats to survival (Purvis et al. 2000, p. 328). Phylogenetic analysis can estimate extinction risks for clades

(Purvis et al. 2000) as well as for biogeographic areas of endemism (Gudde et al. 2013) by assessing if extinction rates are higher than a random expectation based on predicted losses of phylogenetic diversity (Faith 1992; Faller et al. 2008).

Fundamentally, however, inventorial methods are just that: an inventory of the biodiversity present in a given habitat or locality. Thus, species delimitation and biodiversity discovery is an essential component of conservation genetics, especially given the increasingly recognized presence of morphologically cryptic organismal variation (Jorger and Schrodler 2013; Fiser et al. 2018; Struck et al. 2018). Here, in particular, Kimura's theory applies in an unexpected sense. To identify cryptic biodiversity, investigators are increasingly deploying DNA “barcodes,” which are typically short stretches of mitochondrial DNA. Neutral theory predicts that the evolution of monophyly will be four times slower in nuclear than mitochondrial genes due to their differing  $N_e$ , and accordingly, mitochondrial loci will coalesce more rapidly than nuclear loci (Moore 1995; Palumbi et al. 2007). Although coalescence rates can differ between sexually and asexually reproducing organisms, the use of mitochondrial barcodes for detecting recent historical events and for surveying large numbers of taxa with limited sequence data still represents a gold standard for rapid biodiversity inventories. Indeed, it is this direct connection between neutral theory and coalescence that bestows the power of coalescent methods to simultaneously

interpret past events and predict future outcomes—a power that is further enhanced by the increasing flood of unlinked single nucleotide polymorphisms (SNPs) facilitated by the genomics revolution.

### “Functional” Conservation Genetics

Functional (or mechanistic) approaches to conservation genetics largely consist of characterizing functionally important variation related to organismal development and evolutionary potential. Adaptive divergence depends upon genome-wide variation, both within and among populations, created by the interaction of gene flow and local adaptation (Funk et al. 2012). We are able to predict changes in genetic diversity over time (Pauls et al. 2013), and failure to accommodate this in populations of conservation concern can lead to elevated extinction risks. Global extinction rates can now be predicted using quantitative estimates of the future trajectories of biodiversity. These trajectories inform models from which we can predict changes in drivers of extinction risk such as range shifts, habitat loss, and community collapse (Moss et al. 2010). In turn, these inferences can inform policy, protecting not only a population or species but whole ecosystems and their socioeconomic roles. And though in the long term, loss of heterozygosity may be slightly overestimated by standard measures of  $N_e$ , neutral theory will be a powerful predictor according to the “precautionary principle” (Montgomery et al. 2000, p. 42).

Taking adaptive loci into consideration in the management of a population or species protects significant genetic variation that may have substantial ecological and societal benefits (Prince et al. 2017). Many of these features directly relate to Kimura’s (1955) observation that loss of heterozygosity is expected to be inversely related to  $N_e$ . Thus, given that population size is a major determinant of genetic variation, estimates of  $N_e$  are fundamental to determining the present and future health of that population. Avise (2010) further defines the functional role of conservation genomics as differentiating between genetic variation that is neutral versus variation that is adaptive, such as MHC loci in mammals or self-incompatibility loci in plants. Needless to say, such distinctions would be meaningless in the absence of Kimura’s theory given that the observations of exceptionally high or low  $F_{ST}$  values or high ratios of nonsynonymous to synonymous substitutions can only be comprehended against the background of neutral or nearly neutral loci.

### “Process-Driven” Conservation Genetics

The genomics era has enabled conservation geneticists to produce SNP data sets of a magnitude that was unimaginable a decade ago. These data, combined with insights from neutral theory, have accelerated the development of coalescent-based historical demographic analysis in much the same way that the first amino acid sequence data led Kimura (1968) to his neutral theory of molecular evolution. Here, we focus on the development and application of coalescent-based approaches, drawing special attention to the direct links among neutral theory, the coalescent, and conservation

genetics. Among approaches for inferring demographic history, those based on coalescent theory have arguably seen the most progress as a consequence of the explosion of genome-scale data. Not only do they allow for characterization of population histories, they contribute to the development of informed hypotheses regarding potential outcomes. One of the many useful features of genomic coalescent-based methods for conservation geneticists is that small sample sizes often suffice, and powerful analyses can even be applied to single individuals (e.g., PSMC; Li and Durbin 2011). Coalescent-based methods open windows into the past, and thus, the population processes that have yielded the demographic parameters that can be detected in the present (Montano 2016).

The standard coalescent model provides an expectation of the genealogical histories of a population under a constant mutation rate and effective population size. There are additional mathematically simplifying assumptions about the absence of population structure, migration, recombination, and selection. These simplifying assumptions therefore provide a statistical framework under which deviations from the standard coalescent null model can be used to infer changes in demographic processes such as effective population size. Moreover, extensions to the standard coalescent model can be used to infer gene flow between populations. Two categories of coalescent-based methods for demographic inference stand out for having seen much recent development—methods based on the site frequency spectrum (Excoffier et al. 2013; Liu and Fu 2015; Boitard et al. 2016) and methods that explicitly make use of linkage information, for example, by modeling recombination along the genome using Hidden Markov Models (Li and Durbin 2011; Schiffels and Durbin 2014).

Thus, the behavior of neutral alleles is governed by fluctuations in population size, and it is coalescent theory that enables the analytical power to estimate these changes through time (Griffiths and Tavaré 1994; Moore 1995). This window into the demographic past has uniquely enabled conservation geneticists to distinguish between populations that are connected via gene flow from those that are isolated (Wakeley 1996). Testing for population structure using molecular methods is particularly relevant for taxa of conservation concern. These lineages are often comprised of populations with highly fragmented distributions, and it may not be clear to what extent gaps in distribution are recent and anthropogenic rather than the consequence of natural historical processes (Yoder et al. 2016). Coalescent methods excel in detailed investigations of population structure that go beyond the mere inference of structure, extending to the use of isolation-with-migration models that coestimate divergence times and migration rates along with population sizes of ancestral and extant populations. In the end, theoretical advances have empowered the field of conservation genetics to detect the critical and actionable dangers of population subdivision and lost genetic connectivity, as well as the risks of low genetic diversity and inbreeding depression (Shafer et al. 2015).

While many coalescent-based methods, such as PSMC (Li and Durbin 2011) and MSMC (Schiffels and Durbin 2014), examine relatively ancient demographic events, several are illuminating for more recent timescales that are especially relevant to conservation (Chan et al. 2006; Ho and Shapiro 2011; Liu and Fu 2015; Boitard et al. 2016; Nunziata and Weisrock 2018). Particularly pertinent is a recent study that used coalescent-based modeling to study the onset of recent rapid population declines (Li et al. 2016). This work used information across >2,700 species and inferred that population sizes of many currently threatened species have started to decline only in the last few hundred years. Although it remains difficult to directly translate estimates of  $N_e$  to census population sizes (Palstra and Fraser 2012; Ellegren and Galtier 2016), our ability to interpret broad trends in  $N_e$  over time, and across multiple populations and taxa, can provide explicit evidence that global biodiversity is at risk, and that this risk is associated with human activity (Li et al. 2016). This, in turn, is the ammunition required to guide and inspire conservation action.

### Neutral Theory Sets the “Conservation Prior”

Neutral theory serves as the null hypothesis for molecular-evolutionary theory, and particularly, for the coalescent process (Wakeley 2003; Duret 2008), yet “the value of neutral theory in conservation has gone unrecognized” (Rosindell et al. 2011, p. 346). Though Rosindell et al. (2011) are referencing Hubbell’s ecological theory in this case, the point also applies to the neutral theory of molecular evolution. In making the connection between Kimura’s theory with Hubbell’s ecological neutral theory, Rosindell et al. (2011) draw special attention to exchangeability as the common conceptual theme. Whether it is genomic loci or individuals within a community, neutrality implies that substituting one allele or individual for another does not impact the evolutionary fitness of either. Thus, when we observe departures from exchangeability, we are alerted to active processes such as human-mediated climate change, deforestation, the introduction of chemical pollutants, and so on (the list is depressingly long), which have disrupted the neutral expectations of population connectivity and gene flow.

The time has arrived for conservation genetics to adopt an explicit hypothesis-driven framework that draws from Kimura’s neutral theory of molecular evolution. Shafer et al. (2015) have described the need for setting “conservation priors” (p. 84) to guide the design of research questions that are actionable and achievable for improving population and species viability. Our expectations for variation at the molecular level as a function of  $N_e$  and migration can be used in the planning of experimental designs of conservation genetic research that use model-based hypothesis testing to detect likely migration corridors (Aguillon et al. 2017), infer population declines that may explain present-day reduced genetic variation (Figueiró et al. 2017), or evaluate the efficacy of restoration programs (Li et al. 2014). These actionable hypotheses can be advanced through a synthesis of inventorial, functional, and especially process-driven conservation

genomics; a synthesis that leverages the power of the many molecular-evolutionary insights that Kimura bestowed on the field 50 years ago.

### References

- Aguillon SM, Fitzpatrick JW, Bowman R, Schoech SJ, Clark AG, Coop G, Chen N. 2017. Deconstructing isolation-by-distance: the genomic consequences of limited dispersal. *PLoS Genet.* 13(8):e1006911–e1006927.
- Avice JC. 2010. Perspective: conservation genetics enters the genomics era. *Conserv Genet.* 11(2):665–669.
- Benestan LM, Ferchaud AL, Hohenlohe PA, Garner BA, Naylor GJP, Baums IB, Schwartz MK, Kelley JL, Luikart G. 2016. Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Mol Ecol.* 25(13):2967–2977.
- Boitard S, Rodriguez W, Jay F, Mona S, Austerlitz F. 2016. Inferring population size history from large samples of genome-wide molecular data – an approximate bayesian computation approach. *PLoS Genet.* 12(3):e1005877.
- Ceballos G, Ehrlich PR, Barnosky AD, Garcia A, Pringle RM, Palmer TM. 2015. Accelerated modern human-induced species losses: entering the sixth mass extinction. *Sci Adv.* 1(5):e1400253.
- Chan YL, Anderson CNK, Hadly EA. 2006. Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *PLoS Genet.* 2(4):e59–460.
- Di Piero EA, Mosca E, González-Martínez SC, Binelli G, Neale DB, La Porta N. 2017. Adaptive variation in natural Alpine populations of Norway spruce (*Picea abies* [L.] Karst) at regional scale: landscape features and altitudinal gradient effects. *For Ecol Manage.* 405:350–359.
- Duret L. 2008. Neutral theory: the null hypothesis of molecular evolution. *Nat Educ.* 1:218.
- Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nat Rev Genet.* 17(7):422–433.
- Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet.* 9(10):e1003905.
- Exposito-Alonso M, Vasseur F, Ding W, Wang G, Burbano HA, Weigel D. 2018. Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*. *Nat Ecol Evol.* 2(2):352–358.
- Faith DP. 1992. Conservation evaluation and phylogenetic diversity. *Biol Conserv.* 61(1):1–10.
- Faller B, Pardi F, Steel M. 2008. Distribution of phylogenetic diversity under random extinction. *J Theor Biol.* 251(2):286–296.
- Fennessy J, Bidon T, Reuss F, Kumar V, Elkan P, Nilsson MA, Vamberger M, Fritz U, Janke A. 2016. Multi-locus analyses reveal four giraffe species instead of one. *Curr Biol.* 26(18):2543–2549.
- Figueiró HV, Li G, Trindade FJ, Assis J, Pais F, Fernandes G, Santos SHD, Hughes GM, Komissarov A, Antunes A, et al. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Sci Adv.* 3:e1700299.
- Fiser C, Robinson CT, Malard F. 2018. Cryptic species as a window into the paradigm shift of the species concept. *Mol Ecol.* 27(3):613–635.
- Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. 2012. Harnessing genomics for delineating conservation units. *Trends Ecol Evol.* 27(9):489–496.
- Gante HF, Matschiner M, Malmström M, Jakobsen KS, Jentoft S, Salzburger W. 2016. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. *Mol Ecol.* 25(24):6143–6161.
- Griffiths RC, Tavaré S. 1994. Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc Lond B Biol Sci.* 344(1310):403–410.
- Gudde RM, Joy JB, Mooers AO. 2013. Imperilled phylogenetic endemism of Malagasy lemuriformes. *Divers Distributions* 19(7):664–675.
- Ho SYW, Shapiro B. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol Ecol Resour.* 11(3):423–434.

- Jorger KM, Schrod M. 2013. How to describe a cryptic species? Practical challenges of molecular taxonomy. *Front Zool.* 10(1):59.
- Kimura M. 1955. Solution of a process of random genetic drift with a continuous model. *Proc Natl Acad Sci U S A.* 41(3):144–150.
- Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* 217(5129):624–626.
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–U484.
- Li HP, Xiang-Yu JG, Dai GY, Gu ZL, Ming C, Yang ZF, Ryder OA, Li WH, Fu YX, Zhang YP. 2016. Large numbers of vertebrates began rapid population decline in the late 19th century. *Proc Natl Acad Sci U S A.* 113(49):14079–14084.
- Li S, Li B, Cheng C, Xiong Z, Liu Q, Lai J, Carey HV, Zhang Q, Zheng H, Wei S, et al. 2014. Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome Biol.* 15(12):557.
- Liu XM, Fu YX. 2015. Exploring population size changes using SNP frequency spectra (vol 47, pg 555, 2015). *Nat Genet.* 47(9):1099–1099.
- Lynch M, O'Hely M. 2001. Captive breeding and the genetic fitness of natural populations. *Conserv Genet.* 2(4):363–378.
- Montano V. 2016. Coalescent inferences in conservation genetics: should the exception become the rule? *Biol Lett.* 12(6):20160211.
- Montgomery ME, Woodworth LM, Nurthen RK, Gilligan DM, Briscoe DA, Frankham R. 2000. Relationships between population size and loss of genetic diversity: comparisons of experimental results with theoretical predictions. *Conserv Genet.* 1:33–43.
- Moore WS. 1995. Inferring phylogenies from mtDNA variation – mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49(4):718–726.
- Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, van Vuuren DP, Carter TR, Emori S, Kainuma M, Kram T, et al. 2010. The next generation of scenarios for climate change research and assessment. *Nature* 463(7282):747–756.
- Nunziata SO, Weisrock DW. 2018. Estimation of contemporary effective population size and population declines using RAD sequence data. *Heredity* 120(3):196–207.
- Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Hedrick PW. 2010. Conservation genetics in transition to conservation genomics. *Trends Genet.* 26(4):177–187.
- Palstra FP, Fraser DJ. 2012. Effective/census population size ratio estimation: a compendium and appraisal. *Ecol Evol.* 2(9):2357–2365.
- Palumbi SR, Cipriano F, Hare MP. 2007. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evol Int J Org Evol.* 55(5):859–868.
- Pauls SU, Nowak C, Balint M, Pfenninger M. 2013. The impact of global climate change on genetic diversity within populations and species. *Mol Ecol.* 22(4):925–946.
- Primmer CR. 2009. From conservation genetics to conservation genomics. *Year Ecol Conserv Biol.* 1162:357–368.
- Prince DJ, O'Rourke SM, Thompson TQ, Ali OA, Lyman HS, Saglam IK, Hotaling TJ, Spidle AP, Miller MR. 2017. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Sci Adv.* 3(8):e1603198.
- Purvis A, Agapow PM, Gittleman JL, Mace GM. 2000. Nonrandom extinction and the loss of evolutionary history. *Science* 288(5464):328–330.
- Rosindell J, Hubbell SP, Etienne RS. 2011. The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol Evol.* 26(7):340–348.
- Russello MA, Waterhouse MD, Etter PD, Johnson EA. 2015. From promise to practice: pairing non-invasive sampling with genomics in conservation. *PeerJ* 3:e1106.
- Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat Genet.* 46(8):919–925.
- Shafer ABA, Wolf JBW, Alves PC, Bergstrom L, Bruford MW, Brannstrom I, Colling G, Dalen L, De Meester L, Ekblom R, et al. 2015. Genomics and the challenging translation into conservation practice. *Trends Ecol Evol.* 30(2):78–87.
- Steiner CC, Putnam AS, Hoeck PEA, Ryder OA. 2013. Conservation genomics of threatened animal species. *Annu Rev Anim Biosci.* 1(1):261–281.
- Struck TH, Feder JL, Bendiksbj M, Birkeland S, Cerca J, Gusarov VI, Kistenich S, Larsson KH, Liow LH, Nowak MD, et al. 2018. Finding evolutionary processes hidden in cryptic species. *Trends Ecol Evol.* 33(3):153–163.
- Wakeley J. 1996. Distinguishing migration from isolation using the variance of pairwise differences. *Theor Popul Biol.* 49(3):369–386.
- Wakeley J. 2003. Inferences about the structure and history of populations: coalescents and intraspecific phylogeography. In: Singh RS, Uyenoyama MK, editors. *The evolution of population biology.* Cambridge: Cambridge University Press. p. 193–215.
- Yoder AD, Campbell CR, Blanco MB, dos Reis M, Ganzhorn JU, Goodman SM, Hunnicutt KE, Larsen PA, Kappeler PM, Rasoliarison RM, et al. 2016. Geogenetic patterns in mouse lemurs (genus *Microcebus*) reveal the ghosts of Madagascar's forests past. *Proc Natl Acad Sci U S A.* 113(29):8049–8056.