



Emerging Frontiers in the Study of Molecular Evolution

David A. Liberles¹ · Belinda Chang² · Kerry Geiler-Samerotte³ · Aaron Goldman⁴ · Jody Hey¹ · Betül Kaçar⁵ · Michelle Meyer⁶ · William Murphy⁷ · David Posada⁸ · Andrew Storfer⁹

Published online: 14 February 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

A collection of the editors of *Journal of Molecular Evolution* have gotten together to pose a set of key challenges and future directions for the field of molecular evolution. Topics include challenges and new directions in prebiotic chemistry and the RNA world, reconstruction of early cellular genomes and proteins, macromolecular and functional evolution, evolutionary cell biology, genome evolution, molecular evolutionary ecology, viral phylodynamics, theoretical population genomics, somatic cell molecular evolution, and directed evolution. While our effort is not meant to be exhaustive, it reflects research questions and problems in the field of molecular evolution that are exciting to our editors.

Keywords Prebiotic evolution · Comparative genomics · Evolutionary cell biology · Molecular evolutionary ecology · Somatic evolution · Directed evolution

Introduction

Recently, with a new Editor-in-Chief and expanded Editorial Board, a set of changes to *Journal of Molecular Evolution* has begun (Liberles 2019). The journal was founded by Emile Zuckerkandl and has a long history of studies in a diverse array of topics from phylogenetics to protein evolution, origin of life, and evolution of the genetic code. Over the history of the journal, each Editor-in-Chief and appointed editorial board has emphasized different areas while also maintaining current publication trajectories in chemical and abiotic evolution.

The *Journal of Molecular Evolution* now aims to broaden its reach into evolutionary genomics while also recapturing the tradition of publication in molecular phylogenetics, modeling, and theory. With this in mind, new editorial board members were invited to highlight particular areas of molecular evolution that they find particularly compelling. While this effort is not meant to be systematic, or exclusive of areas not discussed, it is meant as an indication of scientific directions that members of the editorial board see as novel and emerging sub-disciplines. Further, while this editorial is not a call for manuscripts, the hope is that this communication will establish *Journal of Molecular Evolution* as a home for such research areas. This view from a collection of our

✉ David A. Liberles
daliberles@temple.edu

¹ Department of Biology and Center for Computational Genetics and Genomics, Temple University, Philadelphia, PA 19122, USA

² Department of Ecology and Evolutionary Biology and Department of Cell and Systems Biology, University of Toronto, 25 Harbord Street, Toronto, ON M5S 3G5, Canada

³ Center for Mechanisms of Evolution, School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

⁴ Department of Biology, Oberlin College and Conservatory, K123 Science Center, 119 Woodland Street, Oberlin, OH 44074, USA

⁵ Department of Molecular and Cell Biology, University of Arizona, Tucson, AZ 85721, USA

⁶ Department of Biology, Boston College, Chestnut Hill, MA 02467, USA

⁷ Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA

⁸ Biomedical Research Center (CINBIO), University of Vigo, Vigo, Spain

⁹ School of Biological Sciences, Washington State University, Pullman, WA 99164, USA

editors is ultimately meant to spur discussion about the field of molecular evolutionary biology as a whole.

Prebiotic Evolution and the RNA World (Bottom up)

The *Journal of Molecular Evolution* has been a traditional venue for publications on the origin of life. Specifically, the journal published foundational studies that contribute to our understanding of a potential RNA world. The RNA world hypothesis in its simplest form states that life evolved from a replicating system of RNAs that served both as genetic carriers of heritable information and as the functional molecules encoded by those genetic carriers (Gilbert 1986). Though the functional range of natural RNAs is narrow, especially with respect to catalysis, *in vitro* selection studies produced catalysts that increase the plausibility that an RNA world scenario preceded cellular life (e.g., Lohse and Szostak 1996; Eklund and Bartel 1996; Lau and Unrau 2009).

In vitro selection of nucleic acids has not only yielded RNA molecules important for multiple applications that will be discussed below (Filonov et al. 2014; Svendsen and Jaffrey 2016; Autour et al. 2016, 2018), but demonstrated what plausible RNA catalyzed RNA or DNA polymerization might have looked like in an RNA world scenario (Horning and Joyce 2016; Samanta and Joyce 2017; Attwater et al. 2018). At the same time, non-enzymatic RNA polymerization (Prywes et al. 2016; Zhang et al. 2017; Hänle and Richert 2018), and the role of crowding (Saha et al. 2018) and encapsulation (Bansho et al. 2016; Matsumura et al. 2016) are also becoming increasingly important factors for understanding plausible scenarios for chemical evolution and RNA replication at the origins of life. Furthermore, as high-throughput sequencing continues to fall in cost, RNA is re-emerging as an experimental model to explore evolutionary concepts such as the fitness landscape and epistasis (Pressman et al. 2017, 2019; Bendixsen et al. 2017).

Recent attempts to reconcile the RNA world concept with other considerations regarding the origin of life have produced a much more complex view of this potential stage in early evolution. Particularly, experimental studies have shown that under certain early Earth conditions, new catalytic RNA functions can be discovered, while the efficiency of known RNA catalysts can be enhanced (Hsiao et al. 2013; Popović et al. 2015). From a theoretical perspective, some have argued that any RNA world metabolism would have relied on prebiotic organic compounds produced by the geochemical setting of life's origin (Goldman et al. 2016) and most dramatically, an RNA world may have co-evolved with prebiotic peptides and a rudimentary translation system (di Giulio 1997; Bowman et al. 2015; but also see Poole et al.

2015). This more complex and nuanced view of a potential RNA world presents an important challenge for future experimental and theoretical work on early evolution.

Early Evolutionary History (Top Down)

The ever-growing understanding of abiotic organic chemistry and synthetic evolutionary biology described above can be a powerful tool for understanding the origin of life because it affords researchers the ability to test a broad range of potential origin of life scenarios. But it is also ahistoric insofar as it can yield insight into how life *may have* originated, but not how it *did* in fact originate from a historical perspective (Pross and Pascal 2013). A parallel approach uses phylogenetic analyses of modern genes, genomes, and proteomes across the tree of life to understand early evolution from a historical perspective. One significant target of early evolution studies is the most recent common ancestor of all extant organisms, usually referred to as the Last Universal Common Ancestor (LUCA) (Becerra et al. 2007; Goldman et al. 2013).

Ever since genomes became available over a sufficiently representative taxonomic range, researchers sought to identify gene families, protein families, protein domains, and protein structures that may have originated at or before the time of the LUCA (e.g., Harris et al. 2003; Mirkin et al. 2003; Delaye et al. 2005; Yang et al. 2005; Ranea et al. 2006; Wang et al. 2007; Weiss et al. 2016). Though the results of these studies sometimes disagree in their particulars (Becerra et al. 2007; Goldman et al. 2013), they portray a LUCA that had a complete translation system similar to those we see in extant organisms (Harris et al. 2003; Goldman et al. 2010; Fournier et al. 2011) and a complex metabolic networks composed of protein enzymes (Braakman and Smith, 2012; Goldman et al. 2012, 2016; Weiss et al. 2016). LUCA also likely had a DNA genome (Forterre 2002; Goldman and Landweber 2012; Poole et al. 2014) and cell membrane (Martin and Russell 2003; Peretó et al. 2004), although these features are less certain since many proteins that support the DNA genome are not homologous between Bacteria and Archaea. Further, archaeal phospholipids have a different structure than bacterial and eukaryotic phospholipids. Even so, LUCA appears to represent a population of organisms that may have had a level of molecular and physiological complexity not too different from some modern organisms. Why we do not see a branch on the universal tree until life had evolved to such a high degree of complexity remains an important and open question.

LUCA was the last common ancestor of all organisms, but not the last common ancestor of all genes. The number of independent gene inventions giving rise to extant genes that predated the first DNA genome remains an open question.

A small number of known gene duplications that took place prior to the last universal common ancestor can give some insight into evolutionary history before the time of LUCA. These universal paralogs were originally used to root the tree of life. The tree of life has no species outgroup, but because each paralog makes its own gene tree that resembles the tree of life, the other paralog can be used to root it (Gogarten and Taiz 1992; Gribaldo and Cammarano 1998). More recently, these universal paralogs have been used to understand evolutionary transitions prior to LUCA. For example, the final steps in the expansion of the canonical genetic code were elucidated by performing ancestral sequence reconstruction on universally paralogous families of aminoacyl-tRNA-synthetase enzymes (Fournier et al. 2011; Fournier and Alm 2015). Molecular evolution prior to LUCA is a burgeoning field that represents a cutting edge in the study of early evolution (Wolf and Koonin 2007).

The pairing of ancestral sequence reconstruction with molecular laboratory techniques has become another powerful tool in understanding early evolutionary history because it allows researchers to study proposed resurrected ancient proteins in the laboratory (Chang and Donoghue 2000). Early examples of this approach resurrected possible translation elongation factor protein, EF-Tu, from the bacterial ancestor to infer that they functioned at an optimal temperature of 55–65 °C (Gaucher et al. 2003, 2008). The same approach was more recently used to infer the evolutionary stability of protein structure within a thioredoxin family from the bacterial, archaeal, and archaeal-eukaryotic common ancestors to the present (Ingles-Prieto et al. 2013). It can also be used to suggest aspects of the ecology and physiology of animals long vanished from the earth, as with investigations of nocturnality in early mammalian lineages (Bickelmann et al. 2015; Liu et al. 2019). The field of ancestral protein resurrection has been further enhanced by the ability to replace a protein with ancestral versions within a living cell (Kacar and Gaucher 2012; Kacar et al. 2017a, b). The transformation of cells with genes encoding the putative ancestral versions of proteins promises to shed further light on the nature of molecular functions encoded in the genomes of early organisms including the last universal common ancestor, and recently, to resurrect ancient biogeochemical signatures (Kacar et al. 2017c; Garcia and Kacar 2019). The evolutionary transitions that occurred by the time the last universal common ancestor emerged include some of the most consequential in all of evolutionary history, shaping the internal structure and physiology of all organisms (Becerra et al. 2007; Goldman et al. 2013), and making life capable of speciation and ecological dispersal (Cantine and Fournier 2018).

Evolution of Genes and Proteins

From the evolution of LUCA to the evolution of extant cellular (and viral) genomes, phylogenetic pipelines have been established to understand gene relationships, selection, and protein functional evolution (Anisimova and Liberles 2012; Anisimova et al. 2013). Tests for selection (Kosiol and Anisimova 2019) and the relationship between protein structure and function over evolutionary time (Liberles et al. 2012; Chi and Liberles 2016) have recently been reviewed elsewhere. Understanding the importance of structural constraint in dictating sequence constraint has been a focus in protein evolution (see for example, Grahn et al. 2011). Early stage models typically treated folding as a global property, but a more localized view of folding stability constraints may dramatically change our understanding. Further, the excess amino acid changes due to positive selection and the reduction of amino acid substitution due to clearly defined folding and functional interactions are not well understood mechanistically. In this context, the missing amino acid substitution due to “negative design” associated with both folding and binding amino acids that would fit within a structure but lead to a fold transition by enabling a more energetically favorable conformation to emerge if substituted in the folding sense (Noivirt-Brik et al. 2009) can not easily be detected. From the perspective of inter-molecular binding specificity, this would result in selective pressures to not bind to potential binding partners where the interaction would be deleterious with amino acid substitutions that would still enable a favorable interaction with the native partner (Liberles et al. 2011; Yang et al. 2012). Such missing substitution due to the “negative design” side of folding and binding specificity can probably be estimated statistically, but identifying the cause of it is a more daunting challenge, especially for current computational methods.

Protein structure is an intermediate between the genotype and the phenotype (function) of a protein. However, protein structure appears to be more highly conserved than either protein coding sequence or protein function. For example, when amino acid sequence divergence is compared to structure divergence between the same sets of proteins, a considerable amount of sequence difference is usually required to produce any appreciable difference in structure (Chothia and Lesk 1986; Illergård et al. 2009). Furthermore, families of proteins that share a common structure often evolve a range of different functions (Furnham et al. 2012). One explanation for the high level of conservation observed in protein structures as compared to protein sequence or protein function is that, there are a limited number of stable and biologically useful protein folds and that these are hard to discover through evolutionary processes. Correspondingly, many sequences can yield such folds, which can in turn be

harnessed to perform many different chemical interactions and transformations. This many-to-few-to-many relationship is an important part of the genotype–phenotype map.

One mode of understanding the link between genotype and phenotype is through evolutionary synthetic biology and experimental evolution. Methodological advances, including deep mutational scanning, have combined sequence data and modeling to better understand the rules of evolutionary processes (see for example Doud et al. 2015). This new understanding ultimately can lead us back to computational biology and predicting new genotype–phenotype relationships. While traditional models of statistical genetics are designed to fit data without the ability to extrapolate, more mechanistic models may have this potential and are a growing area, integrating across layers of biological organization (see for example, Loewe 2016; Lind et al. 2019). This will be described below in more detail.

How Basic Properties of Cells Influence Molecular Evolution

Toward the aim of integrating molecular biology with evolutionary biology, the past five years have seen growing enthusiasm for the idea that the structure and function of basic molecular building blocks (e.g., genomes, proteins, regulatory networks and cells) have a profound influence on evolutionary processes. For example, several studies show how the requirement for globular proteins and RNAs to fold into three-dimensional structures can limit the evolutionary trajectories by which they access new functions or optimize existing ones (Canale et al. 2018; Kurahashi et al. 2018; Pressman et al. 2019). Other studies demonstrate how physical constraints on cell size (Farhadifar et al. 2015) or energetic constraints on cell metabolism (Scott et al. 2014) lead to potentially generalizable ‘scaling laws’ that may have pervasive effects on the evolution of diverse organisms. Preceding this recent enthusiasm is a long history of studies focusing on how generic features of cell systems can drive or constrain evolutionary processes, with important parts of that history unfolding in the *Journal of Molecular Evolution*, as reviewed below.

Early studies found puzzling patterns in the evolutionary rates of different nucleotides or genes, leading to discoveries about how these patterns are generated by the way replication machinery, translation machinery and other cellular machines operate (Crick 1966; *J Mol Biol*; Mazin 1976; Kimura 1980; Sharp and Li 1986; Drummond and Wilke 2008; Shahmoradi et al. 2014). For example, the observation that nucleotides in the third codon position vary more than others is driven by the fact that binding of the cognate tRNA is looser in that codon position (Crick 1966). Others observed and searched for mechanistic explanations as to

why some codons are used more than others to specify particular amino acids (Elton et al. 1976; Berger 1978), again discovering intriguing patterns that could not be understood without considering basic properties of cell systems. Eventually, the observation that highly expressed genes are more biased in their codon usage (Bennetzen and Hall 1982; Sharp and Li 1986) was made clearer by understanding the costs cells encounter when producing highly abundant proteins (Drummond and Wilke 2008). Other patterns of codon bias, including those that distinguish tissue-specific genes, for example, remain to be fully understood (Supek 2016).

A few pivotal papers published in the *Journal of Molecular Evolution* transformed diverse observations into general hypotheses about how generic features of cellular systems influence the way evolution unfolds (Zuckerandl 1997; Stoltzfus 1999). One prominent hypothesis that emerged relates to how ubiquitous errors in DNA replication and transmission can create redundancies (e.g., duplicate genes or duplicate pathways) that promote complexity, innovation, and diversity (Stoltzfus 1999; Force et al. 1999).

A second influential hypothesis asserts that the mere fact that genes and proteins physically interact inside of cells can also promote complexity and innovation (Stoltzfus 1999; Zuckerandl 1997). For example, a protein complex may expand because a mutation that destabilizes a necessary interaction can be compensated by recruitment of another protein that re-stabilizes the complex (Jarvis et al. 1989; Zuckerandl 1994, 1997). A recent high-throughput study confirms that complexity (e.g., the number of proteins in a complex) can increase through processes driven by physical interactions between proteins (Diss et al. 2017). A related hypothesis pertains to the idea that interactions among mutations can open or close evolutionary doors (Zuckerandl 1997). For example, studies of protein and tRNA reveal how mutations that destabilize folding are counterbalanced by those that stabilize it, resulting in entrenchment of some mutations (i.e., they are no longer reversible) as well as the possibility of previously forbidden mutations (Huynen 1996). Study of this topic has recently exploded in large part due to new technologies that allow generation and analysis of many mutants (Shah et al. 2015; Starr et al. 2018; Otwinowski et al. 2018; Kurahashi et al. 2018).

In summary, recent work focusing on how the structure and function of molecular building blocks influences evolutionary outcomes stems from a rich history of studies. This enthusiasm has been further fueled by influential review papers (Zuckerandl 1997; Stoltzfus 1999) and most recently by reviews urging deeper consideration of how higher-level cellular features that have historically received little attention (e.g., organelle structure, energetic costs of metabolism) impact evolutionary processes (Lynch et al. 2014; Phillips and Bowerman 2015; Titus and Goodson 2018). Modern high-throughput phenotyping and genome-editing

techniques including DNA barcoding, CRISPR, single-cell microscopy, and RNA-seq have vastly improved our ability to investigate molecular-level features of cells (Kinney and McCandlish 2019), thus enabling more comprehensive investigations of how these features influence molecular evolution. The *Journal of Molecular Evolution* is committed to continuing its tradition of publishing articles in this area and encourages such submissions.

Genome Evolution

Lineage-specific genome content and architecture are shaped by a collection of population genetic and life history traits. Lynch (2007, 2008) identified effective population size as a modulator of the effectiveness of selection as a key parameter in driving differences in gene number and content, as well as in genome structure across species. It has become clear that the nature of the genotype–phenotype map gives rise to many genotypic solutions to a given phenotypic outcome and this has emerged as an important feature of the evolutionary landscape at the genomic level as well. From the genome of the tunicate, *Oikopleura dioica*, to the nature of gene function in glycolysis, there are many examples of surprising variability in genotypic structure (Denoeud et al. 2010; Orlenko et al. 2016). As the catalog of whole-genome sequences grows across the tree of life from phenotypically diverse species, making genotype–phenotype connections will become more commonplace and comparatively powerful. For example, whole-genome comparisons between species with regressive morphologies (e.g., naked mole rats, cetaceans) show that they carry large suites of inactivating mutations that provide genetic signatures revealing the regulatory architecture of complex, adaptive transitions to new life histories (Huelsmann et al. 2019), with many of these phenotypes serving as naturally occurring mimics of human disease (Emerling et al. 2017). Because of the complexity of the genotype–phenotype space, it is a natural extension that at a sequence level, genomic observations are far from a sampling of an evolutionary equilibrium and additional new mappings are expected to be identified as data continues to accumulate (see for example Povolotskaya and Kondrashov 2010).

In the last decade, it has become apparent that genomes harbor numerous signatures of discordant genealogies (Bravo et al. 2019). This variation is due to complex interactions between natural selection, hybridization, recombination, and effective population size (Hobolth et al. 2007, 2011; Schumer et al. 2018; Martin et al. 2019; Li et al. 2019). We are only in the infancy of discovering the full variation encrypted within the genomes of living organisms, and require new methods to analyze whole-genome data in the context of unique local genomic architectures and modes

of genetic transmission. New approaches that consider the phylogenomic structuring of gene histories along and between chromosomes, and their interaction with recombination rates, natural selection, and demography, will be useful for reliably inferring phylogenetic histories and the role of gene flow in obscuring ancient phylogenetic structure.

New long-read sequencing technologies are finally beginning to open up the “dark matter” of the genome, allowing sequencing of long, repetitive gene families and satellite repeats that were not previously possible. Many of these repetitive elements are known to play roles in disease susceptibility and a variety of other phenotypes, so having complete telomere–telomere sequences (e.g., Miga et al. 2019) will provide unparalleled opportunities for comparative genomics and making genotype–phenotype correlations in non-model organisms. One area that will benefit greatly in this regard is the analysis of gene family evolution. Numerous studies in the literature make biological inferences about adaptation from gene loss and gain events in large multicopy gene families. However, gene counts of segmentally duplicated regions in draft genome assemblies are prone to error and incomplete gene models lead to erroneous biological inferences (Denton et al. 2014). Improved genome assemblies, such as through novel trio-binning approaches (Koren et al. 2018; Rice et al. 2019), will push the field forward so that we can better connect copy number evolution changes to phenotypic innovations. Models for understanding these evolutionary dynamics in a tree reconciliation framework are also an important direction (Konrad et al. 2011; Yohe et al. 2019).

In functionally annotating genomes, enrichment analysis using GO terms or the KEGG Database of pathways have become common. The next level of analysis will involve more computational assessment of gene functions. Clustering of positively selected, differentially expressed genes, or retained duplicates in a pathway or functional category can happen for different reasons. Many studies lack a phylogenetic null model that considers mutational opportunity or the notion that mutation can itself be biased. Further, compensatory covariation (epistasis that is evolutionarily neutral) and directional selection may be more difficult to differentiate than is commonly appreciated. There is a functional way forward, as simple models from biophysical chemistry enable us to relate pathway function, protein concentration, and binding (and catalytic for enzymes) activities of coding sequences in mutation–selection frameworks. This is one potential alternative as a mechanistic modeling framework to more empirical approaches.

Molecular Evolutionary Ecology

With the decreasing cost and increased ease of generating genomic-scale data for non-model organisms, molecular evolutionary ecology has undergone somewhat of a new renaissance period over the past decade. Genomic-scale data, ranging from thousands of SNPs, to hundreds of molecular sequences to whole transcriptomes and genomes have resulted in markedly improved resolution to, for example, detect loci under selection, resolve phylogenies, and study speciation and hybridization. Here, each of these areas of study will be discussed, with a concluding section on future directions.

Detecting Loci Under Selection: Population and Landscape Genomics

One research area that has burgeoned in the genomics age is the search for loci underlying local adaptation (Hoban et al. 2016). Originally, common garden experiments and/or field-based reciprocal transplant experiments were used to document whether populations are locally adapted. Most commonly, local adaptation was inferred if individuals from a population had higher fitness (correlates) in their home environment than an environment away from their natal habitat. More recently, an approach for determining the molecular underpinnings of local adaptation emerged in the analytical frameworks of population genomics (Luikart et al. 2003) and landscape genomics (Joost et al. 2007). The main premise for both lines of inquiry is that, by analyzing a large number of loci, some allele frequencies or genetic distances will be correlated with variation in abiotic (or biotic) variables (Luikart et al. 2003; Joost et al. 2007). Accordingly, sampling occurs in different parts of a species' geographic range that vary in the environmental characteristic of interest, such as rainfall or altitude. Two major analytical frameworks were developed to test for such patterns: outlier detection methods (Foll and Gaggiotti 2008) and genotype-environment association (GEA) analyses (Coop et al. 2010; Rellstab et al. 2015). Briefly, outlier detection methods generate a distribution of locus-specific genetic distance values (such as F_{ST}) and then conduct a statistical test for outlier loci; loci with the highest genetic distance values are indicative of positive selection, and loci with the lowest values indicate they are under purifying selection (Luikart et al. 2003). GEA methods test for correlations between allele frequencies and environmental variables (Coop et al. 2010; Rellstab et al. 2015; Hoban et al. 2016). Eventually, it will become possible to bring more mechanistic approaches that are being developed in molecular evolution into molecular ecology as well.

In examining currently applied methods in molecular ecology, numerous analytical methods were developed to conduct outlier analyses and GEAs, and several simulation studies have followed. Some general lessons can be taken from these studies. One major consideration is the background demography of the species and populations under study. For example, different analytical methods provide different power whether a population has recently expanded (e.g., in the case of an invasive species) or has contracted (e.g., in the case of a species of conservation interest) (deVillemereuil et al. 2014; Lotterhos and Whitlock 2014, 2015). A second insight is that there is always going to be a top X% (with X being the desired cutoff for what is being considered as significant) of loci; that is, with a large number of statistical analyses, a number of loci will always come out as significant. The analytical frameworks all have ways to computationally account for multiple testing and false discovery rates, but the rate of false positive loci still remains high under various methods (deVillemereuil et al. 2014; Lotterhos and Whitlock 2014, 2015). Thirdly, most phenotypic traits under selection have polygenic underpinnings and even the best single locus studies (e.g., a GWAS for human height; Yengo et al. 2018) only explain roughly 10% of the phenotypic variance in a trait. This "missing heritability" (Manolio et al. 2009) means that loci discovered in a population genomics framework likely only explain a small proportion of the adaptive genetic variation in a locally adapted trait. Two other important caveats are related to the fact that most of the landscape genomics studies conducted in non-model species involve analyzing anonymous loci (e.g., SNPs generated by RAD-seq; Lowry et al. 2016). That is, most SNPs determined to be under selection are often not found in a gene or regulatory region, but rather in proximity to one. As such, a "moving window" approach can be used to search for genes within the range of linkage disequilibrium of the candidate SNP. If the species under study has small linkage blocks, however, the true allele under selection may often be missed (Lowry et al. 2016). In general, caution should be used with random marker-based studies of species with small linkage groups because even a fairly large number of SNPs (tens of thousands) may only cover a small portion of the genome (Lowry et al. 2016). Despite these caveats, population and landscape genomics studies have yielded invaluable new information regarding population delineation, conservation and management units, and many candidate loci under selection that have enhanced our understanding of the mechanistic basis for local adaptation (Andrews et al. 2016; Hohenlohe et al. 2018).

From the identification of candidate loci under selection with current methods, establishing a functional role remains a challenge. Transcriptomic sequencing can be conducted without a reference genome, and differential expression of a candidate locus in different populations can be a way to

verify putative function for transcription-based phenotypes. While in its early stages of application to non-model organisms, CRISPR can be used to modify any gene, and thereby test putative function. Such studies may be hard to conduct in vivo, but CRISPR could be conducted in vitro in cultured cell lines. Additionally, more attention could be paid to the influence of biotic factors on local adaptation. To date, most landscape genetics studies have focused on abiotic environmental factors, such as altitude or temperature. However, emerging infectious diseases, or other species, such as key predators or prey can greatly influence patterns of local adaptation. Take, for example, devil facial tumor disease (DFTD) a deadly, transmissible cancer of Tasmanian devils that has caused widespread population declines (McCallum et al. 2009). A landscape genomics study showed that DFTD resulted in a decrease in the strength of local adaptation to abiotic factors, such as precipitation, after the disease arrived (Fraik et al. 2019). To that end, population and landscape genomics studies can move more toward studying biotic interactions among species, referred to as a "landscape community genomics" approach (Hand et al. 2015).

The key idea behind landscape community genomics studies is to meld studies of the effects of abiotic landscape characteristics on the spatial arrangement of populations, with the influence of biotic community interactions to test how ecological dynamics affect genomic variation and gene flow. It has long been recognized that community-level interactions among species can drive evolutionary genetic processes, such as population genetic structure (i.e., "community genetics"; Antonovics 2003; Collins 2003). For example, a study of steelhead trout showed that the genotypes of their trematode parasite resulted in a more accurate assignment of trout to their source population than the trout genotypes themselves (Criscione et al. 2006). Indeed, considering the influence of competition, predation or co-evolution in addition to the spatial arrangement of populations can provide new insights into the evolutionary processes that shape species' distributions (Hand et al. 2015). Explicit models for species interactions in communities that interface with metagenomic and ultimately full genomic data are a futuristic area (Aldebert and Stouffer, 2018; Shoemaker et al. 2019).

Speciation and Hybridization

The availability of genomic-scale data has also greatly improved our ability to study the processes of hybridization and speciation. For a long time, evolutionary biologists were interested in identifying the genes that contribute to reproductive isolation and speciation, or so-called "speciation genes" (Orr et al. 2004). However, it was not until the past decade that scientists began to unravel the effect size

of genes that contribute to reproductive isolation (Nosil and Schluter 2011). For example, across several species of *Drosophila*, approximately 18 genes underpinned intrinsic post-mating isolation (Coyne and Orr 2004).

Ecological speciation, or speciation without geographic isolation has also been a major focus of recent diversification studies (Schluter 2009; Nosil and Schluter 2011). An example is a study of hawthorn maggots that went through a phenological host shift to feed on apple. A selection experiment showed that the phenological host shift entailed genome-wide divergence patterns similar to that observed in natural populations (Egan et al. 2015). In general, understanding the speciation process in the face of gene flow (Feder et al. 2012) has garnered widespread interest, and genomic tools, such as the generation of large numbers of anonymous genome-wide markers, allow for empirical tests of model predictions. Further, genome-wide marker sets allow testing of which parts of the genome are in the process of generating inter-specific divergence via maintenance of reproductive isolation and, conversely, which parts are homogenized by gene flow.

Genomics and next-generation sequencing have also advanced studies of hybridization. For example, the collared and pied flycatchers naturally hybridize, but researchers discovered approximately 50 divergence islands—regions of the genome with about 50× the genetic differentiation of the background and that complex repeat structures appear to drive divergence of the two species (Ellegren et al. 2012). Researchers can now test the proportion of the genome that is introgressed from each of the parental species in a hybrid zone (Gompert and Buerkle 2011; Parchman et al. 2013). A recent study showed that there was shared introgression across two different hybrid zones of spotted and collared towhees, suggesting consistency in areas of the genome affected by gene flow (Kingston et al. 2017). Future genomic studies of hybridization can investigate the joint divergence between nucleotide sequences and transcriptomes, leading to insights in understanding the relative influence between DNA divergence and gene expression levels in maintaining and/or destabilizing hybrid zones. Further, we may be able to better appreciate the genomic basis of reproductive isolation, speciation and hybridization as our understanding of the function of structural genome variation improves, such as the relationship between gene copy number and phenotype.

From Phylogenomics to Phylodynamics

Phylodynamics is an application of phylogenomics to study the evolution of whole parasite genomes, usually those of viruses (Holmes and Grenfell 2009). For example, phylodynamics analyses of HIV showed that the first introduction

of HIV-1 into the new world was most likely in Haiti, with the subsequent US introduction from Haiti in 1969, 12 years earlier than previously thought (Gilbert et al. 2007). A more recent phylodynamic analysis of the major African Ebola outbreak from 2013–2016 showed that the epidemic arose from a single spillover infection in Guinea due to the high genetic similarity of virus genomes sampled early in the epidemic (Gire et al. 2014; Holmes et al. 2016). Despite this early genetic similarity among isolates, the ebola strain named “EBOV Makona” spread to Sierra Leone and Liberia, which then diversified into separate, largely independently evolving clusters. One possibility is to expand phylodynamic studies to parasites other than RNA viruses, although associated analyses may be challenging computationally. Phylogenomics studies can also be applied to understand the evolution of virulence by studying the evolutionary dynamics of cross-species transmission, the associated changes in virulence during host switches, and the genomic basis underlying these changes (Geoghegan and Holmes 2018).

The Changing Role of Theory in Population Genomics

Journal of Molecular Evolution has a long tradition of publishing population genetic research, going back at least to some of the foundational papers in the development of the neutral theory (Kimura and Ohta 1971) and the nearly neutral theory (Ohta 1972). Today in an era of genomics, the once theory-heavy field of population genetics has become increasingly data-driven, and the population genomics of 2020 and beyond can expect to see the growing use of genome-scale data sets. Richard Lewontin famously wrote, some 45 years ago, of the transition of population genetics, from a theory-laden to a data-swamped field (Lewontin 1974). In the case of that particular data-swamping, the theoreticians eventually caught up (e.g. Kingman 1982, Charlesworth et al. 1993, Gillespie 2000), but now it has happened again with genome-level data. The enormous information content in population-genomic data sets drives much of the current research on genetic mapping, on the study of natural selection, and on demographic inference—to mention just three long-standing and still big areas of research. Many researchers who work with quantitative population genetic models, or would like to do so, have found that the scale of the data and the challenges of applying theory on such scales have transformed them into statisticians. This is not a bad thing—as the potential for discovery, and the scope of those discoveries can be great with such large data; for all that, we are still doing model-based statistics. But it remains to be seen how theoreticians can respond to the opportunities and challenges of such large data. Will the future of mathematical and computational work in population genomics be

dominated by the development of new inference technologies (i.e., statistics), as seems likely, given current trends? Will new advances in theory, and kinds of theory emerge complementarily to new inference with existing theory to help us gain a greater understanding of the processes driving the patterns we find in these vast data sets? It is clear that current theory needs expansion in multiple directions to deal accurately with selection in changing and large population sizes or with high mutation rates, to give examples, and that such theory would be welcome to those building a new molecular population genetic understanding of species.

Somatic Molecular Evolution

To date, our understanding of molecular evolution has meant *germline* molecular evolution. However, molecular changes also occur within multicellular organisms, so an individual's cells evolve during their lifetime, generating *somatic* molecular evolution. All individuals are genetic mosaics to a different extent, but this has been largely unexplored with the exception of plants (Antolin and Strobeck 1985). Plants can even pass somatic mutations to their progeny, which sometimes confers adaptive advantages (Simberloff and Leppanen 2019).

In recent years, next-generation sequencing has been fundamental to disentangle somatic evolution at different levels, including genomes, methylomes, or transcriptomes (Posada 2015). Most studies of somatic evolution focus on cancer, for which numerous evolutionary studies, often still descriptive, exist about adaptation, population structure, mutational process and divergence (Williams et al. 2018; Martincorena et al. 2018a, b; Ling et al. 2015; Zhao et al. 2016; Alexandrov et al. 2013; Jiang et al. 2016; Sun et al. 2017; Alves et al. 2019). Not surprisingly, the neutral selection debate has also made its presence at the somatic level and it is still ongoing (Williams et al. 2016; Tarabichi et al. 2018).

More recently, a number of studies have tried to understand how cells evolve in healthy tissues, mostly in humans, including skin, blood, colon, liver, esophagus, or brain (Lopez-Garcia et al. 2010; Lodato et al. 2015; Ma et al. 2015; Martincorena et al. 2015; Blokzijl et al. 2016; Martincorena et al. 2018a, b; Su et al. 2018; Lee-Six et al. 2018; Yokoyama et al. 2019). Such studies show that normal cells also compete for space and resources, and that large clonal expansions can occur within a healthy tissue, often favored by strong positive selection. Understanding how somatic mutations accrue with time, or why mutational rates change among tissues might be essential to understand aging and related chronic diseases of aging, such as diabetes, heart disease, or neurological disorders. Nevertheless, interesting examples also exist outside the human body, generating an understanding of adult

development from a single cell and the accumulation of mutations in the soma (Behjati et al. 2014; Schmid-Siebert et al. 2017; Alemany et al. 2018; Olsen et al. 2019).

Indeed, the growth of single-cell genomics (Macaulay and Voet 2014; Gawad et al. 2016; Tanay and Regev 2017; Baslan and Hicks 2017) and transcriptomics (Stegle et al. 2015) has been fundamental for this endeavor, and is not difficult to predict that it will continue to fuel the study of in somatic molecular evolution, in an intimate relationship with development, aging, and disease (Marioni and Arendt 2017).

Disentangling the molecular evolution of cells in humans and other organisms, addressing questions about cell selection and competition, adaptation, interaction with the microenvironment, diversification, mutational processes, genetic drift, phylogeography, population dynamics or phylogenetics, among many others aspects is futuristic. Upcoming studies will address not just empirical questions, as the studies referred above, but also methodological (Alves et al. 2017; Dou et al. 2018; Singer et al. 2018) and theoretical issues (Nowak et al. 2003; Spencer et al. 2006; Frank, 2010; Cannataro and Townsend, 2018) as well.

Somatic cell evolution can also unite with germline evolution in cases where somatic cells speciate into single-cell eukaryotic organisms. This can be viewed as happening unproductively in most somatic cell cancers. However, there are a few cases where transmissible cancers have emerged from multicellular organisms that persist over evolutionary time, including in canines (Baez-Ortega et al. 2019), twice in Tasmanian Devils (Patchett et al. 2019), and in bivalves (Metzger et al. 2016; Yonemitsu et al. 2019). This is a process of interest to both evolutionary and medical biologists, as well as to conservation biologists.

Evolution as a Tool: Directed Evolution

In 2018, the Nobel Prize in Chemistry was awarded to Frances Arnold, George Smith, and Gregory Winter for their work in applied protein evolution. Touching on both chemical evolution and on the evolution of genes and proteins as key journal topics, this Nobel Prize illustrates the importance of understanding evolutionary principles to not only understanding the function of evolved biomacromolecules but also toward engineering applications.

One driver of recent advances in directed evolution is the technology to effectively create/synthesize libraries with specific diversity constraints. Site saturation mutagenesis in particular has been an effective strategy for allowing both nature and designed protein scaffolds to perform a range of non-biological chemistries including: metathesis (Jeschek et al. 2016), enantioselective organic borylation (Kan et al. 2017), carbon-silicon bond formation (Kan et al. 2016), and

C-H amination (Prier et al. 2017). While such studies do not mimic natural evolutionary processes, the products of these experiments inform us that some limitations of biological catalysts are a result of natural selection rather than inherent biophysical barriers.

A second driving factor in the evolution of protein catalysts is the development and widespread application of approaches that enable higher throughput screening or selection. Rare fitness peaks may be reached through ultra-high-throughput screening. In particular, microfluidic droplets that act as individual microreactors, and continuous evolution systems that circumvent the need for discreet selection rounds, are enabling the creation of enzymes with altered biological function. Microfluidic droplets enable encapsulation of individual cells with reagents that allow fluorescence-activated droplet sorting (FADS) assays for enzyme function utilizing fluorogenic substrates (Obexer et al. 2017), or approaches such as compartmentalized self-replication (CSR), wherein altered DNA polymerases expressed by a cell encapsulated with reagents for DNA replication evolve properties such as isothermal replication (Milligan et al. 2018), a proof-reading reverse transcriptase (Ellefson et al. 2016), and polymerization of the nucleotide analog, alpha-L-threofuranosyl nucleic acid (TNA) (Larsen et al. 2016). Continuous evolution systems that link phage replication within a reservoir of *E. coli* have been used to evolve DNA-binding specificity (Brödel et al. 2016), protease specificity (Packer et al. 2017), and aminoacyl-tRNA synthetases (Bryson et al. 2017) among other protein properties. Although these technologies were originally developed over a decade ago (Ghadessy et al. 2001; Esvelt et al. 2011), recent applications point toward the widespread adaptation of such technology to meet the desires of a growing synthetic biology community.

An example of the impact of progressive screening/selection approaches toward a specific goal is the engineering of *SpCas9*, a CRISPR-Cas9 nuclease from *Streptococcus pyogenes*. *SpCas9* was originally engineered to expand allowed recognition sequences (PAM sequences) using a bacterial selection system with sequential positive and negative selections (Kleinstiver et al. 2015). Yet, for many applications, including widespread use in genome editing, further expansion of accessible sequences and reduction of off-target activity is critical. Thus, additional studies using simultaneous positive and negative selection in *E. coli* (Lee et al. 2018) or positive selection coupled with negative screening in yeast (Casini et al. 2018) enabled greater increases in specificity. Subsequently, this problem was approached using phage-assisted continuous evolution (PACE) wherein a virally encoded catalytically inactive Cas9 variant was tethered to a viral RNA polymerase necessary for phage amplification (Hu et al. 2018); thus guide RNA-dependent gene expression activation was used to link Cas9 DNA-binding

and gene expression in a selection suitable for PACE. The CRISPR-Cas9 example also demonstrates how directed evolution approaches incorporating random mutagenesis processes can be complementary with various structure-based rational engineering approaches (Slaymaker et al. 2016; Chen et al. 2017).

Beyond, the *in vitro* evolution of molecules, the *in vitro* evolution of organisms enables experimental control and replication of key evolutionary variables not controllable in natural settings (for example, known selective pressures and population sizes). The most famous of such experiments is surely the Lenski experiment with *E. coli*, which has now been underway for > 25 years and over 60,000 generations (Lenski and Travisano 1994; Lenski 2017). Although this study was started well before the first bacterial genome was sequenced, inexpensive whole-genome sequencing in the last decade has allowed experimental demonstration of important evolutionary concepts including clonal interference (Maddamsetti et al. 2015), epistasis (Khan et al. 2011; Plucain et al. 2014), and convergence (Blount et al. 2018). This work has also been inspirational for others, and today there are numerous studies that incorporate lab adaptation as a strategy for understanding biological systems, or utilize laboratory adaptation as tool for industrial biotechnology (Remigi et al. 2019; Sandberg et al. 2019). Taking things one step further, inter-specific interactions can be examined *in vitro*, toward the design of *in vitro* ecosystems, with control of the organismal make-up, starting conditions, and ecosystem parameters (Lindemann et al. 2016; D'Souza et al. 2018). As theory has progressed in population genetics, molecular evolution, and molecular evolutionary ecology that gives expectations about allele frequencies, genome sequences, and species distributions over time, experimental tests in controlled settings are now possible and represent an exciting development moving forward.

Other Key Directions and Concluding Thoughts

The continued development of theory linking population-level processes through evolution to molecular-level processes is a critical element of molecular evolution. From theoretical developments, models that join intra-specific processes with inter-specific timescales and those that mechanistically capture sequence variation through the genotype–phenotype map are another area of interest. As a next step, new models lead to new methods and approaches for inference in computational biology. Standard assumptions of site independence, time homogeneity, and processes at equilibrium are made for mathematical and computational ease, but can be sufficiently violated in some biological situations to make inference that rests upon such assumptions questionable. Theory interplays with its implementation in computational methods and its application to new data that is generated with new

experimental methods. All of this promises to be an exciting future for the field, including in the pages of *Journal of Molecular Evolution*.

Acknowledgements We thank John Bracht, David Pollock, and an anonymous reviewer for helpful comments.

References

- Aldebert C, Stouffer DB (2018) Community dynamics and sensitivity to model structure: towards a probabilistic view of process-based model predictions. *J R Soc Interface* 15:20180741. <https://doi.org/10.1098/rsif.2018.0741>
- Alemany A, Florescu M, Baron CS et al (2018) Whole-organism clone tracing using single-cell sequencing. *Nature* 556:108–112
- Alexandrov LB, Nik-Zainal S, Wedge DC et al (2013) Signatures of mutational processes in human cancer. *Nature* 500:415–421
- Alves JM, Prieto T, Posada D (2017) Multiregional tumor trees are not phylogenies. *Trends Cancer* 3:546–550
- Alves JM, Prado-Lopez S, Cameselle-Teijeiro JM, Posada D (2019) Rapid evolution and biogeographic spread in a colorectal cancer. *Nat Commun* 10:5139
- Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet* 17:81–92
- Anisimova M, Liberles DA (2012) Detecting and understanding natural selection. In: Cannarozzi GM, Schneider A (eds) *Codon evolution*. Oxford University Press, Oxford, pp 73–96
- Anisimova M, Liberles DA, Philippe H, Provan J, Pupko T, von Haeseler A (2013) State-of-the-art methodologies dictate new standards for phylogenetic analysis. *BMC Evol Biol* 13:161. <https://doi.org/10.1186/1471-2148-13-161>
- Antolin MF, Strobeck C (1985) The population genetics of somatic mutation in plants. *Am Nat* 126:52–62
- Antonovics J (2003) Toward community genomics. *Ecology* 84:598–601
- Attwater J, Raguram A, Morgunov AS et al (2018) Ribozyme-catalysed RNA synthesis using triplet building blocks. *Elife* 7:2804. <https://doi.org/10.7554/eLife.35255>
- Autour A, Westhof E, Ryckelynck M (2016) iSpinach: a fluorogenic RNA aptamer optimized for *in vitro* applications. *Nucleic Acids Res* 44:2491–2500. <https://doi.org/10.1093/nar/gkw083>
- Autour A, Jeng CYS, Cawte DA et al (2018) Fluorogenic RNA mango aptamers for imaging small non-coding RNAs in mammalian cells. *Nat Commun* 9:656. <https://doi.org/10.1038/s41467-018-02993-8>
- Baez-Ortega A, Gori K, Strakova A et al (2019) Somatic evolution and global expansion of an ancient transmissible cancer lineage. *Science* 365(6452):eaau9923. <https://doi.org/10.1126/science.aau9923>
- Bansho Y, Furubayashi T, Ichihashi N, Yomo T (2016) Host-parasite oscillation dynamics and evolution in a compartmentalized RNA replication system. *Proc Natl Acad Sci USA* 113:4045–4050. <https://doi.org/10.1073/pnas.1524404113>
- Baslan T, Hicks J (2017) Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat Rev Cancer* 17:557–569
- Becerra A, Delaye L, Islas S, Lazcano A (2007) The very early stages of biological evolution and the nature of the last common ancestor of the three major cell domains. *Annu Rev Ecol Evol Syst* 38:361–379. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095825>

- Behjati S, Huch M, van Boxtel R et al (2014) Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* 513:422–425
- Bendixsen DP, Østman B, Hayden EJ (2017) Negative epistasis in experimental RNA fitness landscapes. *J Mol Evol* 85:159–168. <https://doi.org/10.1007/s00239-017-9817-5>
- Bennetzen JL, Hall BD (1982) Codon selection in yeast. *J Biol Chem* 257:3026–3031
- Berger EM (1978) Pattern and chance in the use of the genetic code. *J Mol Evol* 10:319–323
- Bickelmann C, Morrow JM, Du J, Schott RK, van Hazel I, Lim S, Müller J, Chang BS (2015) The molecular origin and evolution of dim-light vision in mammals. *Evolution* 69(11):2995–3003. <https://doi.org/10.1111/evo.12794>
- Blokzijl F, de Ligt J, Jager M et al (2016) Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 538:260–264
- Blount ZD, Lenski RE, Losos JB (2018) Contingency and determinism in evolution: replaying life's tape. *Science* 362:eeam5979
- Bowman JC, Hud NV, Williams LD (2015) The ribosome challenge to the RNA world. *J Mol Evol* 80:1–19
- Braakman R, Smith E (2012) The emergence and early evolution of biological carbon-fixation. *PLoS Comput Biol* 8:e1002455. <https://doi.org/10.1371/journal.pcbi.1002455>
- Bravo GA, Antonelli A, Bacon CD, Bartoszek K, Blom MPK, Huynh S, Jones G, Knowles LL, Lamichhaney S, Marcussen T, Morlon H, Nakhleh LK, Oxelman B, Pfeil B, Schliep A, Wahlberg N, Werneck FP, Wiedenhoeft J, Willows-Munro S, Edwards SV (2019) Embracing heterogeneity: coalescing the tree of life and the future of phylogenomics. *Peer J* 7:e6399. <https://doi.org/10.7717/peerj.6399>
- Brödel AK, Jaramillo A, Isalan M (2016) Engineering orthogonal dual transcription factors for multi-input synthetic promoters. *Nat Commun* 7:13858. <https://doi.org/10.1038/ncomms13858>
- Bryson DI, Fan C, Guo L-T et al (2017) Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat Chem Biol* 13:1253–1260. <https://doi.org/10.1038/nchembio.2474>
- Canale AS, Cote-Hammarlof PA, Flynn JM, Bolon DN (2018) Evolutionary mechanisms studied through protein fitness landscapes. *Curr Opin Struct Biol* 48:141–148
- Cannataro VL, Townsend JP (2018) Neutral theory and the somatic evolution of cancer. *Mol Biol Evol* 35:1308–1315
- Cantine MD, Fournier GP (2018) Environmental adaptation from the origin of life to the last universal common ancestor. *Orig Life Evol Biosph* 48:35–54. <https://doi.org/10.1007/s11084-017-9542-5>
- Casini A, Olivieri M, Petris G et al (2018) A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol* 36:265–271. <https://doi.org/10.1038/nbt.4066>
- Chang BS, Donoghue MJ (2000) Recreating ancestral proteins. *Trends Ecol Evol* 15(3):109–114
- Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular evolution. *Genetics* 134:1289–1303
- Chen JS, Dagdas YS, Kleinstiver BP et al (2017) Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550:407–410. <https://doi.org/10.1038/nature24268>
- Chi PB, Liberles DA (2016) Selection on protein structure, interaction, and sequence. *Protein Sci* 25(7):1168–1178. <https://doi.org/10.1002/pro.2886>
- Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J* 5:823–826. <https://doi.org/10.1002/j.1460-2075.1986.tb04288.x>
- Collins JP (2003) What can we learn from community genetics? *Ecology* 84:574–577
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185:1411–1423
- Coyne JA, Orr A (2004) *Speciation*. Oxford University Press, Oxford, p 545
- Crick FH (1966) Codon–anticodon pairing: the wobble hypothesis. *J Mol Biol* 19:548–555
- Criscione CD, Cooper B, Blouin MS (2006) Parasite genotypes identify source populations of migratory fish more accurately than fish genotypes. *Ecology* 87:823–828
- D'Souza G, Shitov S, Preussger D, Yousif G, Waschina S, Kost C (2018) Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Nat Prod Rep* 35:455–488
- de Villemereuil P, Fricot É, Bazin É, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? *Mol Ecol* 23:2006–2019
- Delaye L, Becerra A, Lazcano A (2005) The last common ancestor: what's in a name? *Orig Life Evol Biosph* 35:537–554. <https://doi.org/10.1007/s11084-005-5760-3>
- Denoeud F, Henriet S, Mungpakdee S et al (2010) Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* 330:1381–1385. <https://doi.org/10.1126/science.1194167>
- Denton RD, Kenyon LJ, Greenwald KR, Gibbs H (2014) Evolutionary basis of mitonuclear discordance between sister species of mole salamanders (*Ambystoma* sp.). *Mol Ecol* 23:2811–2824. <https://doi.org/10.1111/mec.12775>
- di Giulio M (1997) On the RNA world: evidence in favor of an early ribonucleopeptide world. *J Mol Evol* 45:571–578
- Diss G, Gagnon-Arsenault I, Dion-Coté A-M et al (2017) Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. *Science* 355:630–634
- Dou Y, Gold HD, Luquette LJ, Park PJ (2018) Detecting somatic mutations in normal cells. *Trends Genet* 34:545–557
- Doud MB, Ashenberg O, Bloom JD (2015) Site-specific amino acid preferences are mostly conserved in two closely related protein homologs. *Mol Biol Evol* 32:2944–2960. <https://doi.org/10.1093/molbev/msv167>
- Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134:341–352
- Egan SP, Ragland G, Assour L, Powell THQ, Hood GR, Emrich S, Nosil P, Feder JL (2015) Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. *Ecol Lett* 18:817–825
- Ekland EH, Bartel DP (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* 382:373–376
- Ellefsen JW, Gollihar J, Shroff R et al (2016) Synthetic evolutionary origin of a proofreading reverse transcriptase. *Science* 352:1590–1593. <https://doi.org/10.1126/science.aaf5409>
- Ellegren H, Smeds L, Burri R, Olason PI, Backström N et al (2012) The genomic landscape of species divergence in *Ficedula flycatchers*. *Nature* 491:756–760
- Elton RA, Russell GJ, Subak-Sharpe JH (1976) Doublet frequencies and codon weighting in the DNA of *Escherichia coli* and its phages. *J Mol Evol* 8:117–135
- Emerling CA, Widjaja AD, Nguyen NN, Springer MS (2017) Their loss is our gain: regressive evolution in vertebrates provides genomic models for uncovering human disease loci. *J Med Genet* 54(12):787–794. <https://doi.org/10.1136/jmedgenet-2017-104837>
- Esvelt KM, Carlson JC, Liu DR (2011) A system for the continuous directed evolution of biomolecules. *Nature* 472:499–503. <https://doi.org/10.1038/nature09929>

- Farhadifar R, Baer CF, Valfort A-C et al (2015) Scaling, selection, and evolutionary dynamics of the mitotic spindle. *Curr Biol* 25:732–740
- Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends Genet* 28:342–350
- Filonov GS, Moon JD, Svensen N, Jaffrey SR (2014) Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J Am Chem Soc* 136:16299–16308. <https://doi.org/10.1021/ja508478x>
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180:977–993
- Force A, Lynch M, Pickett FB et al (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545
- Forterre P (2002) The origin of DNA genomes and DNA replication proteins. *Curr Opin Microbiol* 5:525–532. [https://doi.org/10.1016/S1369-5274\(02\)00360-0](https://doi.org/10.1016/S1369-5274(02)00360-0)
- Fournier GP, Alm EJ (2015) Ancestral reconstruction of a pre-LUCA aminoacyl-tRNA synthetase ancestor supports the late addition of Trp to the genetic code. *J Mol Evol* 80:171–185. <https://doi.org/10.1007/s00239-015-9672-1>
- Fournier GP, Andam CP, Alm EJ, Gogarten JP (2011) Molecular evolution of aminoacyl tRNA synthetase proteins in the early history of life. *Orig Life Evol Biosph* 41:621–632. <https://doi.org/10.1007/s11084-011-9261-2>
- Fraik AK, Margres M, Epstein B, Jones M, Hendricks S, Schönfeld B, Stahlke A, Hamede R, McCallum HI, Lopez-Contreras E, Kallinen SJ, Hohenlohe PA, Kelley JL, Storfer A (2019) Disease swamps molecular signatures of genetic-environmental associations to abiotic factors in Tasmanian devil (*Sarcophilus harrisii*) populations. *BioRxiv*. <https://doi.org/10.1101/780122>
- Frank SA (2010) Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. *Proc Natl Acad USA* 107(suppl 1):1725–1730
- Furnham N, Sillitoe I, Holliday GL, Cuff AL, Laskowski RA et al (2012) Exploring the evolution of novel enzyme functions within structurally defined protein superfamilies. *PLoS Comput Biol* 8:e1002403. <https://doi.org/10.1371/journal.pcbi.1002403>
- Garcia AK, Kaçar B (2019) How to resurrect ancestral proteins as proxies for ancient biogeochemistry. *Free Radic Biol Med* 140:260–269. <https://doi.org/10.1016/j.freeradbiomed.2019.03.033>
- Gaucher EA, Thomson JM, Burgan MF, Benner SA (2003) Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* 425:285–288. <https://doi.org/10.1038/nature01977>
- Gaucher EA, Govindarajan S, Ganesh OK (2008) Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* 451:704–707. <https://doi.org/10.1038/nature06510>
- Gawad C, Koh W, Quake SR (2016) Single-cell genome sequencing: current state of the science. *Nat Rev Genet* 17:175–188
- Geogheghan JL, Holmes EC (2018) The phylodynamics of evolving virus virulence. *Nat Rev Genet* 19:756–769
- Ghadessy FJ, Ong JL, Holliger P (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proc Natl Acad Sci USA* 98:4552–4557. <https://doi.org/10.1073/pnas.071052198>
- Gilbert W (1986) The RNA world. *Nature* 319:618
- Gilbert MTP, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M (2007) The emergence of HIV/ Aids in the Americas and beyond. *PNAS* 104:18566–18570
- Gillespie JH (2000) Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* 155:909–919
- Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ et al (2014) Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 345:1369–1372
- Gogarten JP, Taiz L (1992) Evolution of proton pumping ATPases: rooting the tree of life. *Photosynth Res* 33:137–146. <https://doi.org/10.1007/BF00039176>
- Goldman AD, Landweber LF (2012) Oxytricha as a modern analog of ancient genome evolution. *Trends Genet* 28:382–388. <https://doi.org/10.1016/j.tig.2012.03.010>
- Goldman AD, Samudrala R, Baross JA (2010) The evolution and functional repertoire of translation proteins following the origin of life. *Biol Direct* 5:15. <https://doi.org/10.1186/1745-6150-5-15>
- Goldman AD, Baross JA, Samudrala R (2012) The enzymatic and metabolic capabilities of early life. *PLoS ONE* 7:e39912. <https://doi.org/10.1371/journal.pone.0039912>
- Goldman AD, Bernhard TM, Dolzhenko E, Landweber LF (2013) LUCApedia: a database for the study of ancient life. *Nucleic Acids Res* 41:1079–1082. <https://doi.org/10.1093/nar/gks1217>
- Goldman AD, Beatty JT, Landweber LF (2016) The TIM Barrel architecture facilitated the early evolution of protein-mediated metabolism. *J Mol Evol* 82:17–26. <https://doi.org/10.1007/s00239-015-9722-8>
- Gompert Z, Buerkle C (2011) Bayesian estimation of genomic clines. *Mol Ecol* 20:2111–2127
- Grahn JA, Nandakumar P, Kubelka J, Liberles DA (2011) Biophysical and structural considerations for protein sequence evolution. *BMC Evol Biol* 11:361. <https://doi.org/10.1186/1471-2148-11-361>
- Gribaldo S, Cammarano P (1998) The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *J Mol Evol* 47:508–516. <https://doi.org/10.1007/PL00006407>
- Hand BK, Lowe WH, Kovach RP, Muhfield CC, Luikart G (2015) Landscape community genomics: understanding eco-evolutionary processes in complex environments. *Trends Ecol Evol* 30:161–168
- Hänle E, Richert C (2018) Enzyme-free replication with two or four bases. *Angew Chem Int Ed* 57:8911–8915. <https://doi.org/10.1002/anie.201803074>
- Harris JK, Kelley ST, Spiegelman GB, Pace NR (2003) The genetic core of the universal ancestor. *Genome Res* 13:407. <https://doi.org/10.1101/gr.652803>
- Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB et al (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *Am Nat* 188:379–397. <https://doi.org/10.1086/688018>
- Hobolth A, Christensen OF, Mailund T, Schierup MH (2007) Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet* 3(2):e7. <https://doi.org/10.1371/journal.pgen.0030007>
- Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T (2011) Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res* 21(3):349–356. <https://doi.org/10.1101/gr.114751.110>
- Hohenlohe PA, Hand BK, Andrews KR, Luikart G (2018) Population genomics provides key insights in ecology and evolution Population genomics. Springer, New York, pp 483–510
- Holmes EC, Grenfell BT (2009) Discovering the phylodynamics of RNA viruses. *PLoS Comput Biol* 5:e1000505
- Holmes EC, Dudas G, Rambaut A, Andersen KG (2016) The evolution of Ebola virus: insights from the 2013–2016 epidemic. *Nature* 538:193–200
- Horning DP, Joyce GF (2016) Amplification of RNA by an RNA polymerase ribozyme. *Proc Natl Acad Sci USA* 113:9786–9791. <https://doi.org/10.1073/pnas.1610103113>

- Hsiao C, Chou I-C, Okafor CD et al (2013) RNA with iron(II) as a cofactor catalyses electron transfer. *Nat Chem* 5:525–528
- Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, Zeina CM, Gao X, Rees HA, Lin Z, Liu DR (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556(7699):57–63. <https://doi.org/10.1038/nature26155>
- Huelsmann M, Hecker N, Springer MS, Gatesy J, Sharma V, Hiller M (2019) Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations. *Sci Adv* 5(9):eaaw6671. <https://doi.org/10.1126/sciadv.aaw6671>
- Huynen MA (1996) Exploring phenotype space through neutral evolution. *J Mol Evol* 43:165–169
- Illergård K, Ardell DH, Elofsson A (2009) Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. *Proteins* 77:499–508. <https://doi.org/10.1002/prot.22458>
- Ingles-Prieto A, Ibarra-Molero B, Delgado-Delgado A, Perez-Jimenez R, Fernandez JM, Gaucher EA, Sanchez-Ruiz JM, Gavira JA (2013) Conservation of protein structure over four billion years. *Structure* 21:1690–1697. <https://doi.org/10.1016/j.str.2013.06.020>
- Jarvis EE, Clark KL, Sprague GF Jr (1989) The yeast transcription activator PRTF, a homolog of the mammalian serum response factor, is encoded by the MCM1 gene. *Genes Dev* 3:936–945
- Jeschek M, Reuter R, Heinisch T et al (2016) Directed evolution of artificial metalloenzymes for in vivo metathesis. *Nature* 537:661–665. <https://doi.org/10.1038/nature19114>
- Jiang Y, Qiu Y, Minn AJ, Zhang NR (2016) Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proc Natl Acad Sci* 113:E5528–E5537
- Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G et al (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol Ecol* 16:3955–3969
- Kacar B, Gaucher EA (2012) Towards the recapitulation of ancient history in the laboratory: combining synthetic biology with experimental evolution. *Artif Life* 13:11–18. <https://doi.org/10.7551/978-0-262-31050-5-ch002>
- Kacar B, Garmendia E, Tuncbag N, Andersson DI, Hughes D (2017a) Functional constraints on replacing an essential gene with its ancient and modern homologs. *MBio* 8:e01276. <https://doi.org/10.1128/mBio.01276-17>
- Kacar B, Ge X, Sanyal S, Gaucher EA (2017b) Experimental evolution of *Escherichia coli* harboring an ancient translation protein. *J Mol Evol* 84:69–84. <https://doi.org/10.1007/s00239-017-9781-0>
- Kacar B, Guy L, Smith E, Baross J (2017c) Resurrecting ancestral genes in bacteria to interpret ancient biosignatures. *Philos Trans A Math Phys Eng Sci*. <https://doi.org/10.1098/rsta.2016.0352>
- Kan SBJ, Lewis RD, Chen K, Arnold FH (2016) Directed evolution of cytochrome c for carbon-silicon bond formation: bringing silicon to life. *Science* 354:1048–1051. <https://doi.org/10.1126/science.aah6219>
- Kan SBJ, Huang X, Gumulya Y et al (2017) Genetically programmed chiral organoborane synthesis. *Nature* 552:132–136. <https://doi.org/10.1038/nature24996>
- Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 322:1193–1196
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120
- Kimura M, Ohta T (1971) On the rate of molecular evolution. *J Mol Evol* 1:1–17
- Kingman JFC (1982) The coalescent. *Stoch Process Appl* 13:235–248
- Kingston SE, Parchman TL, Gompert Z, Buerkle CA, Braun MJ (2017) Heterogeneity and concordance in locus-specific differentiation and introgression between species of towhees. *J Evol Biol* 30:474–485
- Kinney JB, McCandlish DM (2019) Massively parallel assays and quantitative sequence-function relationships. *Annu Rev Genomics Hum Genet* 20:99–127
- Kleinstiver BP, Prew MS, Tsai SQ et al (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523:481–485. <https://doi.org/10.1038/nature14592>
- Konrad A, Teufel AI, Grahnen JA, Liberles DA (2011) Toward a general model for the evolutionary dynamics of gene duplicates. *Genome Biol Evol* 3:1197–1209. <https://doi.org/10.1093/gbe/evr093>
- Koren S, Rhie A, Walenz B et al (2018) De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol* 36:1174–1182. <https://doi.org/10.1038/nbt.4277>
- Kosiol C, Anisimova M (2019) Selection acting on genomes. *Methods Mol Biol* 1910:373–397. https://doi.org/10.1007/978-1-4939-9074-0_12
- Kurahashi R, Sano S, Takano K (2018) Protein evolution is potentially governed by protein stability: directed evolution of an esterase from the hyperthermophilic archaeon *Sulfolobus tokodaii*. *J Mol Evol* 86:283–292
- Larsen AC, Dunn MR, Hatch A et al (2016) A general strategy for expanding polymerase function by droplet microfluidics. *Nat Commun* 7:11235. <https://doi.org/10.1038/ncomms11235>
- Lau MWL, Unrau PJ (2009) A promiscuous ribozyme promotes nucleotide synthesis in addition to ribose chemistry. *Chem Biol* 16:815–825
- Lee JK, Jeong E, Lee J et al (2018) Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun*. <https://doi.org/10.1038/s41467-018-05477-x>
- Lee-Six H, Øbro NF, Shepherd MS et al (2018) Population dynamics of normal human blood inferred from somatic mutations. *Nature* 561:473–478
- Lenski RE (2017) Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J* 11:2181–2194. <https://doi.org/10.1038/ismej.2017.69>
- Lenski RE, Travisano M (1994) Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc Natl Acad Sci USA* 91:6808–6814
- Lewontin RC (1974) The genetic basis of evolutionary change. Columbia University Press, New York
- Li G, Figueiró HV, Eizirik E, Murphy WJ (2019) Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. *Mol Biol Evol* 36(10):2111–2126. <https://doi.org/10.1093/molbev/msz139>
- Liberles DA (2019) A new editorial beginning at journal of molecular evolution. *J Mol Evol* 87:69–71
- Liberles DA, Tisdell MD, Grahnen JA (2011) Binding constraints on the evolution of enzymes and signalling proteins: the important role of negative pleiotropy. *Proc Biol Sci* 278:1930–1935. <https://doi.org/10.1098/rspb.2010.2637>
- Liberles DA, Teichmann SA, Bahar I et al (2012) The interface of protein structure, protein biophysics, and molecular evolution. *Protein Sci* 21:769–785. <https://doi.org/10.1002/pro.2071>
- Lind PA, Libby E, Herzog J, Rainey PB (2019) Predicting mutational routes to new adaptive phenotypes. *Elife* 8:e38822. <https://doi.org/10.7554/eLife.38822>
- Lindemann SR, Bernstein HC, Song HS, Fredrickson JK, Fields MW, Shou W, Johnson DR, Beliaev AS (2016) Engineering microbial consortia for controllable outputs. *ISME J* 10:2077–2084

- Ling S, Hu Z, Yang Z et al (2015) Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc Natl Acad Sci USA* 112:E6496–E6505
- Liu Y, Cui Y, Chi H, Xia Y, Liu H, Rossiter SJ, Zhang S (2019) Scotopic rod vision in tetrapods arose from multiple early adaptive shifts in the rate of retinal release. *Proc Natl Acad Sci U S A* 116(26):12627–12628. <https://doi.org/10.1073/pnas.1900481116>
- Lodato MA, Woodworth MB, Lee S et al (2015) Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* 350:94–98
- Loewe L (2016) Systems in evolutionary systems biology. In: Kliman RM (ed) *The encyclopedia of evolutionary biology*. Oxford Academic Press, Elsevier, pp 297–318
- Lohse PA, Szostak JW (1996) Ribozyme-catalysed amino-acid transfer reactions. *Nature* 381:442–444
- Lopez-Garcia C, Klein AM, Simons BD, Winton DJ (2010) Intestinal stem cell replacement follows a pattern of neutral drift. *Science* 330:822–825
- Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol Ecol* 23:2178–2192
- Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol* 24:1031–1046
- Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A (2016) Breaking RAD: the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol Ecol Resour*. <https://doi.org/10.1111/1755-0998.12635>
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* 4:981–994. <https://doi.org/10.1038/nrg1226>
- Lynch M (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci USA* 104(Suppl 1):8597–8604
- Lynch M (2008) *The origins of genome architecture*. Sinauer Associates Inc., Sunderland
- Lynch M, Field MC, Goodson HV et al (2014) Evolutionary cell biology: two origins, one objective. *Proc Natl Acad Sci USA* 111:16990–16994
- Ma H, Folmes CDL, Wu J et al (2015) Metabolic rescue in pluripotent cells from patients with mtDNA disease. *Nature* 524:234–238
- Macaulay IC, Voet T (2014) Single cell genomics: advances and future perspectives. *PLoS Genet* 10:e1004126
- Maddamssetti R, Lenski RE, Barrick JE (2015) Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics* 200:619–631
- Manolio TA, Collins FS, Cox NJ, Goldstein DB et al (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753
- Marioni JC, Arendt D (2017) How single-cell genomics is changing evolutionary and developmental biology. *Annu Rev Cell Dev Biol* 33:537–553
- Martin W, Russell MJ (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc London Ser B* 358:59–83. <https://doi.org/10.1098/rstb.2002.1183>
- Martin SH, Davey JW, Salazar C, Jiggins CD (2019) Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol* 17(2):e2006288. <https://doi.org/10.1371/journal.pbio.2006288>
- Martincorena I, Roshan A, Gerstung M et al (2015) Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 348:880–886
- Martincorena I, Fowler JC, Wabik A et al (2018a) Somatic mutant clones colonize the human esophagus with age. *Science* 362:911–917
- Martincorena I, Raine KM, Gerstung M et al (2018b) Universal patterns of selection in cancer and somatic tissues. *Cell* 173:1823
- Matsumura S, Kun Á, Ryckelynck M et al (2016) Transient compartmentalization of RNA replicators prevents extinction due to parasites. *Science* 354:1293–1296. <https://doi.org/10.1126/science.aag1582>
- Mazin AL (1976) Evolution of DNA structure: direction, mechanism, rate. *J Mol Evol* 8:211–249
- McCallum H, Jones M, Hawkins C, Hamede R et al (2009) Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. *Ecology* 90:3379–3392
- Metzger MJ, Villalba A, Carballal MJ, Iglesias D, Sherry J, Reinisch C, Muttray AF, Baldwin SA, Goff SP (2016) Widespread transmission of independent cancer lineages within multiple bivalve species. *Nature* 534:705–709
- Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, Brooks S, Howe E, Porubsky D, Logsdon GA, Schneider VA, Potapova T, Wood J, Chow W, Armstrong J, Fredrickson J, Pak E, Tigyi K, Kremitzki M, Markovic C, Maduro V, Dutra A, Bouffard GG, Chang AM, Hansen NF, Thibaud-Nissen F, Schmitt AD, Belton J-M, Selvaraj S, Dennis MJ, Soto DC, Sahasrabudhe R, Kaya G, Quick J, Loman NJ, Holmes N, Loose M, Surti U, Risques RA, Graves Lindsay TA, Fulton R, Hall I, Paten B, Howe K, Timp W, Young A, Mullikin JC, Pevzner PA, Gerton JL, Sullivan BA, Eichler EE, Phillippy AM (2019) Telomere-to-telomere assembly of a complete human X chromosome. *bioRxiv*. <https://doi.org/10.1101/735928>
- Milligan JN, Shroff R, Garry DJ, Ellington AD (2018) Evolution of a thermophilic strand-displacing polymerase using high-temperature isothermal compartmentalized self-replication. *Biochemistry* 57:4607–4619. <https://doi.org/10.1021/acs.biochem.8b00200>
- Mirkin BG, Fenner TI, Galperin MY, Koonin EV (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol Biol* 3:2. <https://doi.org/10.1186/1471-2148-3-2>
- Noivirt-Brik O, Horovitz A, Unger R (2009) Trade-off between positive and negative design of protein stability: from lattice models to real proteins. *PLoS Comput Biol* 5:e1000592. <https://doi.org/10.1371/journal.pcbi.1000592>
- Nosil P, Schluter D (2011) The genes underlying the process of speciation. *Trends Ecol Evol* 26:160–167
- Nowak MA, Michor F, Iwasa Y (2003) The linear process of somatic evolution. *Proc Natl Acad Sci USA* 100:14966–14969
- Obexer R, Godina A, Garrabou X et al (2017) Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat Chem* 9:50–56. <https://doi.org/10.1038/nchem.2596>
- Ohta T (1972) Population size and rate of evolution. *J Mol Evol* 1:305–314
- Olsen KC, Moscoso JA, Levitan DR (2019) Somatic mutation is a function of clone size and depth in *Orbicella* reef-building corals. *Biol Bull* 236:1–12
- Orlenko A, Hermansen RA, Liberles DA (2016) Flux control in glycolysis varies across the tree of life. *J Mol Evol* 82:146–161. <https://doi.org/10.1007/s00239-016-9731-2>
- Orr HA, Masly JP, Presgraves DC (2004) Speciation genes. *Curr Opin Genet Dev* 14:675–679
- Otwinowski J, McCandlish DM, Plotkin JB (2018) Inferring the shape of global epistasis. *Proc Natl Acad Sci USA* 115:E7550–E7558
- Packer MS, Rees HA, Liu DR (2017) Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat Commun* 8:956. <https://doi.org/10.1038/s41467-017-01055-9>

- Parchman TLZ, Gompert MJ, Braun RT, Brumfield DB, McDonald JAC, Uy G, Zhang ED, Jarvis BA, Buerkle SCA (2013) The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Mol Ecol* 22:3304–3317
- Patchett AL, Coorens THH, Darby J et al (2019) Two of a kind: transmissible Schwann cell cancers in the endangered Tasmanian devil (*Sarcophilus harrisii*). *Cell Mol Life Sci*. <https://doi.org/10.1007/s00018-019-03259-2>
- Peretó J, López-García P, Moreira D (2004) Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci* 29:469–477. <https://doi.org/10.1016/j.tibs.2004.07.002>
- Phillips PC, Bowerman B (2015) Cell biology: scaling and the emergence of evolutionary cell biology. *Curr Biol* 25:R223–R225
- Plucaín J, Hindre T, Le Gac M, Tenaillon O, Cruveiller S, Medigue C, Leiby N, Harcombe WR, Marx CJ, Lenski RE, Schneider D (2014) Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. *Science* 343:1366–1367
- Poole AM, Horinouchi N, Catchpole RJ, Si D, Hibi M, Tanaka K, Ogawa J (2014) The case for an early biological origin of DNA. *J Mol Evol* 79:204–212. <https://doi.org/10.1007/s00239-014-9656-6>
- Poole AM, Jeffares DC, Hoepfner MP, Penny D (2015) Does the ribosome challenge our understanding of the RNA world? *J Mol Evol* 82:1–4
- Popović M, Fliss PS, Ditzler MA (2015) In vitro evolution of distinct self-cleaving ribozymes in diverse environments. *Nucleic Acids Res* 43:7070–7082
- Posada D (2015) Cancer molecular evolution. *J Mol Evol* 81:81–83
- Povolotskaya IS, Kondrashov FA (2010) Sequence space and the ongoing expansion of the protein universe. *Nature* 465:922–926. <https://doi.org/10.1038/nature09105>
- Pressman A, Moretti JE, Campbell GW et al (2017) Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences. *Nucleic Acids Res* 45:8167–8179. <https://doi.org/10.1093/nar/gkx540>
- Pressman AD, Liu Z, Janzen E et al (2019) Mapping a systematic ribozyme fitness landscape reveals a frustrated evolutionary network for self-aminoacylating RNA. *J Am Chem Soc* 141:6213–6223. <https://doi.org/10.1021/jacs.8b13298>
- Prier CK, Zhang RK, Buller AR et al (2017) Enantioselective, intermolecular benzylic C-H amination catalysed by an engineered iron-haem enzyme. *Nat Chem* 9:629–634. <https://doi.org/10.1038/nchem.2783>
- Pross A, Pascal R (2013) The origin of life: what we know, what we can know and what we will never know. *Open Biol* 3:120190. <https://doi.org/10.1098/rsob.120190>
- Prywes N, Blain JC, Del Frate F, Szostak JW (2016) Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *Elife* 5:1859. <https://doi.org/10.7554/eLife.17756>
- Ranea JA, Sillero A, Thornton JM, Orengo CA (2006) Protein superfamily evolution and the last universal common ancestor (LUCA). *J Mol Evol* 63:513–525. <https://doi.org/10.1007/s00239-005-0289-7>
- Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Mol Ecol* 24:4348–4370
- Remigi P, Masson-Boivin C, Rocha EPC (2019) Experimental evolution as a tool to investigate natural processes and molecular functions. *Trends Microbiol* 27:623–634
- Rice E, Koren S, Rhie A, Heaton M, Kalbfleisch T, Hardy T, Hackett P, Bickhart D, Rosen B, Ley B, Maurer N, Green R, Phillippy A, Petersen J, Smith T (2019) Chromosome-length haplotigs for yak and cattle from trio binning assembly of an F1 hybrid. *BioRxiv*. <https://doi.org/10.1101/737171>
- Saha R, Verbanic S, Chen IA (2018) Lipid vesicles chaperone an encapsulated RNA aptamer. *Nat Commun*. <https://doi.org/10.1038/s41467-018-04783-8>
- Samanta B, Joyce GF (2017) A reverse transcriptase ribozyme. *Elife* 6:2804. <https://doi.org/10.7554/eLife.31153>
- Sandberg TE, Salazar MJ, Weng LL, Palsson BO, Feist AM (2019) The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab Eng* 56:1
- Schluter D (2009) Evidence for ecological speciation and its alternative. *Science* 323:737–741
- Schmid-Siegert E, Sarkar N, Iseli C et al (2017) Low number of fixed somatic mutations in a long-lived oak tree. *Nat Plants* 3:926–929
- Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, Blazier JC, Sankararaman S, Andolfatto P, Rosenthal GG, Przeworski M (2018) Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* 360(6389):656–660. <https://doi.org/10.1126/science.aar3684>
- Scott M, Klumpp S, Mateescu EM, Hwa T (2014) Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol Syst Biol* 10:747
- Shah P, McCandlish DM, Plotkin JB (2015) Contingency and entrenchment in protein evolution under purifying selection. *Proc Natl Acad Sci USA* 112:E3226–E3235
- Shahmoradi A, Sydykova DK, Spielman SJ et al (2014) Predicting evolutionary site variability from structure in viral proteins: buriedness, packing, flexibility, and design. *J Mol Evol* 79:130–142
- Sharp PM, Li WH (1986) An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol* 24:28–38
- Shoemaker LG, Barner AK, Bittleston LS, Teufel AI (2019) Quantifying the relative importance of competition, predation, and environmental variation for species coexistence. *BiorXiv*. <https://doi.org/10.1101/797704>
- Simberloff D, Leppanen C (2019) Plant somatic mutations in nature conferring insect and herbicide resistance. *Pest Manage Sci* 75:14–17
- Singer J, Kuipers J, Jahn K, Beerenwinkel N (2018) Single-cell mutation identification via phylogenetic inference. *Nat Commun* 9:5144
- Slymaker IM, Gao L, Zetsche B et al (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science* 351:84–88. <https://doi.org/10.1126/science.aad5227>
- Spencer SL, Gerety RA, Pienta KJ, Forrest S (2006) Modeling somatic evolution in tumorigenesis. *PLoS Comput Biol* 2:e108
- Starr TN, Flynn JM, Mishra P et al (2018) Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. *Proc Natl Acad Sci USA* 115:4453–4458
- Stegle O, Teichmann SA, Marioni JC (2015) Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* 16:133–145
- Stoltzfus A (1999) On the possibility of constructive neutral evolution. *J Mol Evol* 49:169–181
- Su T, Grady JP, Afshar S et al (2018) Inherited pathogenic mitochondrial DNA mutations and gastrointestinal stem cell populations. *J Pathol* 246:427–432
- Sun R, Hu Z, Sottoriva A et al (2017) Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nat Genet* 49:1015–1024
- Supek F (2016) The code of silence: widespread associations between synonymous codon biases and gene function. *J Mol Evol* 82:65–73
- Svensen N, Jaffrey SR (2016) Fluorescent RNA aptamers as a tool to study RNA-modifying enzymes. *Cell Chem Biol* 23:415–425. <https://doi.org/10.1016/j.chembiol.2015.11.018>
- Tanay A, Reggev A (2017) Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541:331

- Tarabichi M, Martincorena I, Gerstung M et al (2018) Neutral tumor evolution? *Nat Genet* 50:1630–1633
- Titus MA, Goodson HV (2018) Developing evolutionary cell biology. *Dev Cell* 47:395–396
- Wang M, Yafremava LS, Caetano-Anollés D, Mittenthal JE, Caetano-Anollés G (2007) Reductive evolution of architectural repertoires in proteomes and the birth of the tripartite world. *Genome Res* 17:1572–1585. <https://doi.org/10.1101/gr.6454307>
- Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF (2016) The physiology and habitat of the last universal common ancestor. *Nat Microbiol* 1:16116. <https://doi.org/10.1038/nmicrobiol.2016.116>
- Williams MJ, Werner B, Barnes CP et al (2016) Identification of neutral tumor evolution across cancer types. *Nat Genet* 48:238–244
- Williams MJ, Werner B, Heide T et al (2018) Quantification of subclonal selection in cancer from bulk sequencing data. *Nat Genet* 50:895–903
- Wolf YI, Koonin EV (2007) On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol Direct* 2:14
- Yang S, Doolittle RF, Bourne PE (2005) Phylogeny determined by protein domain content. *Proc Natl Acad Sci USA* 102:373–378. <https://doi.org/10.1073/pnas.0408810102>
- Yang J-R, Liao B-Y, Zhuang S-M, Zhang J (2012) Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc Natl Acad Sci USA* 109:E831–E840
- Yengo L, Sidorenko J, Kempner KE, Zheng Z, Wood AR et al (2018) Meta-analysis of genome-wide association studies for height and body mass index in ~ 700000 individuals of European ancestry. *Hum Mol Genet* 27:3641–3649
- Yohe LR, Liu L, Dávalos LM, Liberles DA (2019) Protocols for the molecular evolutionary analysis of membrane protein gene duplicates. In: Sikosek T (ed) *Computational methods in protein evolution* *Methods in molecular biology*. Humana Press, New York, pp 1–5
- Yokoyama A, Kakiuchi N, Yoshizato T et al (2019) Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* 565:312–317
- Yonemitsu MA, Giersch RM, Polo-Prieto M et al (2019) A single clonal lineage of transmissible cancer identified in two marine mussel species in South America and Europe. *eLife* 8:449
- Zhang W, Tam CP, Walton T et al (2017) Insight into the mechanism of nonenzymatic RNA primer extension from the structure of an RNA-GpppG complex. *Proc Natl Acad Sci USA* 114:7659–7664. <https://doi.org/10.1073/pnas.1704006114>
- Zhao Z-M, Zhao B, Bai Y et al (2016) Early and multiple origins of metastatic lineages within primary tumors. *Proc Natl Acad Sci USA* 113:2140–2145
- Zuckerandl E (1994) Molecular pathways to parallel evolution: I. Gene nexuses and their morphological correlates. *J Mol Evol* 39:661–678
- Zuckerandl E (1997) Neutral and nonneutral mutations: the creative mix—evolution of complexity in gene interaction systems. *J Mol Evol* 44(Suppl 1):S2–8