Unlocking the hidden dimensions of genomic diversity within species

Marina Brasó-Vives#, Diego A. Hartasánchez#, Julien F. Ayroles, Christabel F. Bucao, Mayra Furlan-Magaril, Paco Majic, Luisa F. Pallares, James Phipps-Tan, Thea F. Rogers, Aurora Ruiz-Herrera, Gabriela Santos-Rodriguez, Nikolas Vellnow, Justin J. S. Wilcox, Juan Antonio Rodríguez*

* Corresponding author

# Equal contribution

Author details

Marina Brasó-Vives
- Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
- Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland
- marina.brasovives@unil.ch
- ORCID: 0000-0001-5007-5789

Diego A. Hartasánchez
- Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
- diego.hartasanchezfrenk@unil.ch
- ORCID: 0000-0003-2596-6883

Julien F. Ayroles
- Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey, United States of America
- Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, United States of America
- jayroles@princeton.edu
- ORCID: 0000-0001-8729-0511

Christabel F. Bucao
- Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
- Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland
- christabelfloi.bucao@unil.ch
- ORCID: 0000-0001-6705-0940

Mayra Furlan-Magaril
- Molecular Genetics Department, Institute of Cellular Physiology, National Autonomous University of Mexico, Mexico City, Mexico
- mfurlan@ifc.unam.mx
- ORCID: 0000-0002-6771-6631

Paco Majic
- European Molecular Biology Laboratory, Heidelberg, Germany
- paco.majic@embl.de
- ORCID: 0000-0003-0670-7413

Luisa F. Pallares
- Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany

- luisa.pallares@tuebingen.mpg.de
- ORCID: 0000-0001-6547-1901

James Phipps-Tan
- Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany
- james.tanshengyi@tuebingen.mpg.de

Thea F. Rogers
- Department of Neuroscience and Developmental Biology, Division of Molecular Evolution and Development, University of Vienna, Vienna, Austria
- thea.rogers@univie.ac.at
- ORCID: 0000-0003-0198-0242

Aurora Ruiz-Herrera
- Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain
- Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain
- aurora.ruizherrera@uab.cat
- ORCID: 0000-0003-3868-6151

Gabriela Santos-Rodriguez
- School of Biotechnology & Biomolecular Sciences, University of New South Wales, Sydney, Australia
- RNA Institute, University of New South Wales, Sydney, Australia
- g.santosrodriguez@unsw.edu.au
- ORCID: 0000-0002-9076-0294

Nikolas Vellnow
- TU Dortmund University, Computational Systems Biology, Faculty of Biochemical and Chemical Engineering, Dortmund, Germany
- nikolas.vellnow@tu-dortmund.de
- ORCID: 0000-0002-7054-3268

Justin J. S. Wilcox
- TU Dortmund University, Computational Systems Biology, Faculty of Biochemical and Chemical Engineering, Dortmund, Germany
- justin.wilcox@tu-dortmund.de

Juan Antonio Rodríguez:
- Center for Evolutionary Hologenomics, Globe Institute, University of Copenhagen, Copenhagen, Denmark.
- juan.rodriguez@sund.ku.dk
- ORCID: 0000-0001-5195-315X

Abstract

The missing heritability problem, defined as the failure of genetic variants to explain variance in phenotypes, has been an unsolved issue in genetics for the past two decades. A potential solution to this problem stems from the idea that single nucleotide polymorphisms and copy-number variants, the most commonly studied forms of genomic diversity, do not represent the totality of the information that is transferred from one generation to the next. In this perspective, we offer a glimpse into hidden dimensions of genomic diversity: frequently neglected as sources of variation, and only recently starting to be perceived as targets for natural selection. We begin by acknowledging the existence of disregarded genomic matter and sequence composition affecting the mutational process. We explore how selection can act on gene expression variability and offer insight into the mechanisms behind transcriptional regulation. Finally, we recognize the importance of modifications in the 3D genome architecture and other large-scale genomic changes in generating phenotypic diversity. Zooming out through these dimensions of the genome, we discuss recent findings and future avenues of research within each one of them. Judging by the pervasiveness and diverse nature of sources of genomic diversity, we foresee that the study of genomic variation within species is yet to see another burst. Showcasing these hidden dimensions, we offer a broad perspective on the complexity behind the genotype-to-phenotype map, which is crucial for understanding the action of natural selection in populations.

Statement of significance

Genomic diversity stands as the cornerstone and ultimate target of natural selection. During the last couple of decades, huge advances have been made in characterizing genomic diversity within a population. Mainly single nucleotide polymorphisms (SNPs) but also copy-number variants (CNVs) have emerged as the primary sources of genomic variation and have attracted practically all the attention, yet, they only represent two facets of the genomic landscape. Genomes exhibit remarkable variability in numerous alternative dimensions: epigenetic, 3D chromatin conformation and gene expression variation are known to be heritable, have impact on phenotype and, thus, to be targets for natural selection. These alternative dimensions have been traditionally used as complements to SNPs and CNVs, but the vast realm of their population-specific diversity has often been ignored. Acknowledging this diversity opens new perspectives for understanding the forces of natural selection and molecular targets, disease mechanisms, and genetic interactions with the environment.

Main text

Beyond SNPs and CNVs

Genomic variation is the main source of heritable phenotypic variability upon which natural selection acts. The development of sequencing technologies in the last decades has opened the gates for the detailed exploration of this diversity, revealing numerous instances in which detectable genomic changes explain the phenotypic diversity observed in species (e.g., Jeffares et al. 2015). During this time, mainly single nucleotide polymorphisms (SNPs) but also copy number variants (CNVs) have attracted most of the attention in the field. They are both essential to assess the level of within-population variation and are known to be under selection in extant populations (Saitou & Gokcumen 2020). However, their prominence in the field may stem more from methodological convenience than biological significance. Importantly, there are several other forms of genomic diversity that are also targeted by selection. Although the latter have been theoretically proposed for many years, experimental evidence of selectable variation beyond SNPs and CNVs is relatively recent.

In parallel, one of the great ambitions of the field in the last century has been to relate genotype to phenotype. Despite the significant development of quantitative genetics and the infinitesimal model, there is still a fundamental gap in our understanding of the *genotype-to-phenotype map*. This is evidenced by the *missing heritability problem* (Pallares 2019), and the failure of genome-wide association studies to reveal causal variants for phenotypes (Matthews & Turkheimer 2022). New models of inheritance, such as the omnigenic model (Boyle et al. 2017), models incorporating the spatiotemporal organization of the genome (Misteli 2007), or models considering the genome as a generative model of the organism (Mitchell & Cheney 2024) are expanding our understanding of the genotype-to-phenotype map, challenging the way Mendelian paradigms have influenced it.

New hidden dimensions

Inspired by the work presented in the symposium “Unlocking the hidden dimensions of genomic diversity within species”, during the 2024 SMBE Meeting in Puerto Vallarta, Mexico, this perspective will transit through alternative/hidden/unexplored genome dimensions harboring potentially selectable within-species variation. We will discuss a spectrum of hidden genomic diversity, zooming out from neglected sequence variation to genomic rewiring through large scale rearrangements, suggesting that there are likely many more yet to be discovered.

Disregarded genome

During the last 20 years we have seen a plethora of genomes for thousands of species being published (Lewin et al. 2022). Despite claims of completion, it is evident that many published genomes are still missing substantial parts of genetic material. For instance, the comparison between sequencing-independent methods with sequencing-based methods to estimate genome size suggests that a substantial part of the genetic material is not yet included in reference genomes (Elliott & Gregory 2015). Much of this *disregarded genome* stems from unassembled parts of already known chromosomes, such as from repetitive regions that are difficult to assemble (Hartasánchez et al. 2018), especially with short-read sequencing technologies. Recent work on chickens has, for example, found that thousands of genes are missing from the assemblies of birds due to their locations in repetitive content and near chromosome ends (Li et al. 2022). An even larger source of unaccounted-for genomic matter are supernumerary chromosomes like the germline-restricted chromosome in

songbirds (Borodin et al. 2022) or extrachromosomal DNA (Paulsen et al. 2018). Such forms of variation are particularly interesting due to their tissue-specific consequences and ephemeral existence.

Mutational potential

Even within extensively studied genomic sequences, there are overlooked aspects such as their *mutational potential*. The mutational potential of any given sequence is defined by both its *mutational neighborhood* (i.e., the number of sequential point mutations that separate any two sequences) and its *mutational propensities* (i.e., the likelihood of a mutational step being taken over others). The latter could come, for example, in the shape of molecular biases (Cano et al. 2023) or heritability biases due to internal conflicts (Majic & Payne 2023). Mutational neighborhoods influence the evolutionary potential of populations because they define, together with genotype-to-phenotype maps, the accessibility to alternative phenotypes and, therefore, the mutational potential for adaptation (Martin et al. 2024). For example, the specific architecture of the genetic code is such that certain codons are more mutationally volatile than others, even if they encode for the same amino acid. There is, consequently, a different evolutionary potential in each protein-coding sequence that is dependent on its mutational neighborhood. This implies that if the genetic code had a different architecture, neighborhoods would change together with the evolutionary potential of proteins (Rozhoňová et al. 2024; Tsour et al. 2024). Furthermore, mutational neighborhoods can also influence evolution whenever they may be revealed by non-heritable mutations during phenotypic development, through, for example, transcriptional, translational or somatic mutations (Whitehead et al. 2008; Yanagida et al. 2015; Majic et al. 2022).

Gene expression level variability

The immense majority of studies dealing with the action of natural selection upon mutations focus on the mean of any phenotypic trait. In contrast, the evolution of *phenotypic variability*, or, how far individuals tend to be from the mean, has received less attention. In unicellular organisms, research on bet-hedging, gene expression noise, and robustness show that phenotypic variability, though constrained under fixed environments, helps populations find fitter phenotypes faster under unpredictable or rapidly fluctuating environments (Wagner 2012; LaBoone & Assis 2024). It is less clear how variability evolves in multicellular eukaryotes. The few available studies focus primarily on variability in gene transcript levels and its correlates. Common findings include genes with lower transcriptional variability having higher essentiality, lower genetic diversity, fewer protein-protein interactions and tighter regulatory architectures (Sigalova et al. 2020; Wolf et al. 2023).

While these correlations point towards the evolutionary potential of *gene expression variability*, many core evolutionary factors remain underexplored. For example, what is the role of cell type- or tissue-dependent mean expression in constraining gene-specific variability? Given that multicellular eukaryotes have functionally specialized cell and tissue types sharing a quasi-identical genomic background, we can expect some pleiotropic constraints on the regulation of gene expression across cells or tissues. For instance, brain-biased genes have recently been found to have low transcriptional variability across organs in ray-finned fishes, suggesting that such genes may have evolved under stricter regulatory architectures compared to other genes (Bucao et al. 2024), whereas high transcriptional variability can allow for organ-specific fine-tuning, for example, in floral organs (Hartasánchez et al. 2023). Other pressing questions relate to the environmental impact on transcriptional variability and the nature of the genetic variation regulating such variability. Both are

critical to (i) address how transcriptional variability can be selected for and (ii) explicitly account for genotype-by-environment interactions, phenomena proving to be more abundant than commonly appreciated (Tautz et al. 2020). While difficult to study in the past, these questions may now be addressed using cost-effective techniques that allow the collection of RNA-DNAseq for thousands of individuals from populations experimentally evolved under multiple controlled environments (Pallares et al. 2020, 2023). Thus, while research has traditionally focused on mean transcript levels, transcriptional variability studies contribute complementary information towards understanding how complex traits are regulated and evolve.

Transcriptional diversity

The previous section highlights how the amount of phenotypic variability, specifically gene expression level variability, can be a target of selection. Moreover, the transcriptional process itself can lead to variation in transcriptional products, which can be targets of selection. *Transcriptional diversity* is the result of both the existence of different transcriptional products (i.e., mRNAs, and several types of non-coding RNAs), and of multiple transcription-related mechanisms. The best known of such mechanisms are: (i) *alternative promoter usage* that provides different transcription start sites (Reyes & Huber 2018), (ii) *alternative splicing* that produces different transcripts from one gene (Wright et al. 2022), and (iii) *alternative polyadenylation* that generates mRNA 3' end diversity (Mitschka & Mayr 2022). In addition to these three, many other transcription-related mechanisms can be found throughout the tree of life (Ji et al. 2011; Santos-Rodriguez et al. 2021). The wide variety of transcript types and transcription-related mechanisms leads to enormous transcriptional diversity, which varies between tissues and cell types and has been shown to contribute to evolution (Barbosa-Morais et al. 2012; Necsulea et al. 2014; Santos-Rodriguez et al. 2021). One instance of tissue-specific evolutionary conservation of transcription diversity is found in erythroid cells (Ji et al. 2011). For example, RNA Polymerase II pausing seems to be a shared hallmark of erythropoiesis from chickens to humans (Murphy et al. 2021; Penagos-Puig et al. 2023).

Previous research, restricted by short-read sequencing, could only describe a small amount of the transcriptional diversity. With long-read technologies, we can now assess full-length transcripts and RNA modifications, leading to new studies on the phenotypic impact of transcriptome diversity in evolution. We foresee extensive future work, focusing on how changes in regulatory mechanisms that shape transcript levels and diversity contribute to transcriptional programs and species evolution.

3D genome architecture

Not only is the linear DNA sequence the source of hidden dimensions; the folding of the 30-nanometer DNA fiber is also crucial in defining gene expression programs. Chromosomes are organized into structures that span different genomic scales, from chromosome territories, chromatin compartments, *topologically associating domains* (TADs), to chromatin loops (Rowley et al. 2017). Genome organization is dynamic and displays cell type-specific patterns paired with gene expression. Understanding the diverse patterns of chromatin structure across different taxa is essential to uncover the mechanisms behind genome evolution. This diversity is reflected at the chromosome occupancy level, where at least four patterns of nuclear architecture including (i) chromosome territories, (ii) clustering of both centromeres and telomeres, (iii) centromere clustering, and (iv) telomere clustering have been described (Hoencamp et al. 2021; Álvarez-González, Arias-Sardá, et al. 2022; Álvarez-González & Ruiz-Herrera 2024).

At the fine scale, chromatin forms long-range interactions that connect enhancers and promoters. The specificity of chromatin looping between regulatory elements is partly achieved by insulating these interactions within TADs, which are delimited by structural proteins such as CTCF and cohesins. TAD boundaries' integrity is important for transcriptional regulation (Nora et al. 2017; Rao et al. 2017), although there are differences across taxa. CTCF originated in the bilaterian ancestor (Heger et al. 2012) and is a well-documented organizer of genomic compartments in mammals (Dixon et al. 2012), zebrafish (Franke et al. 2021) and *Xenopu*s (Niu et al. 2021). However, in *Drosophila* and *Anopheles* the primary means of chromatin organization is into A/B compartments, with loop extrusion via CTCF being limited to just a few loci such as the *Hox* cluster (Kaushal et al. 2021), although these species possess other architectural proteins besides CTCF (Pauli et al. 2016). Currently, the impact of changes in genome size and structure on genome topology and gene regulation remains underexplored (Mota-Gómez et al. 2022; Rogers & Simakov 2023; Álvarez-González & Ruiz-Herrera 2024). The presence of TAD-like domains outside bilaterians raises intriguing questions about the homology of these structures and the conservation of communication between cis-regulatory elements and promoters across diverse topological frameworks (Acemel & Lupiáñez 2023). Future research should focus on overcoming methodological challenges to better evaluate TAD conservation across clades, as current results vary significantly depending on the approaches used (Eres & Gilad 2020). Furthermore, a systematic evaluation of population-level variation in 3D genome structures, and how they are affected by environmental variables, is still lacking. Both dimensions (across clades and across populations) will provide new insights into our understanding of genome evolution and the underlying mechanisms of gene regulation.

Large-scale genomic changes

Zooming out further, we encounter large-scale genomic processes, such as chromosomal fusions, fissions, translocations, and inversions, which can result in affected cis- and trans-regulation (Harewood & Fraser 2014; Vara et al. 2021; Vara & Ruiz-Herrera 2022). However, the precise mechanisms by which these chromosomal changes shape regulatory evolution and gene expression remain unclear.

Analyses of conserved genomic regions, such as *microsynteny* (Irimia et al. 2012; Albertin et al. 2022), *macrosynteny* (Álvarez-González, Arias-Sardá, et al. 2022; Yu et al. 2024) and *spatiosynteny* (Clarence et al. 2023), along with *evolutionary breakpoints* (Vara et al. 2021; Álvarez-González, Burden, et al. 2022), reveal how intrachromosomal territories and whole chromosomes, evolve under diverging selective constraints, altering the regulatory landscape. Whole-genome duplications add another layer of complexity, allowing for gene family and regulatory element expansions that can undergo neofunctionalization or enhance existing regulatory pathways (Cañestro et al. 2013; Van de Peer et al. 2017). Genome expansion can also be driven by segmental duplications (Brasó-Vives et al. 2022; Vollger et al. 2022) or repetitive element proliferation in non-coding regions, introducing novel regulatory elements to the genome (Sundaram & Wysocka 2020; Albertin et al. 2022).

Conversely, genome contraction through the purging of non-essential sequences, such as redundant repeats and pseudogenes, may lead to the reorganization of existing genes and regulatory elements, causing them to cluster closer together, rewiring expression patterns (Rogers & Simakov 2023). A recent model integrates evolutionary genome reshuffling with DNA damage response mechanisms and the dynamic spatial genome organization of germ cells (Álvarez-González, Burden, et al. 2022; Álvarez-González & Ruiz-Herrera 2024). In this context, chromosomal interactions resulting from

large-scale genomic processes may rewire and/or attenuate gene networks in the germline, fostering evolutionary novelty.

Concluding remarks

Alternative facets of genomic variation offer promising perspectives for improving our understanding of the genotype-to-phenotype map. We argue that all sources of genomic variation influence each other in a quasi-hierarchical manner, and could be thought as layers of such a map. These layers extend across developmental and life stages and environments, implying that a full understanding of how genetic variation propagates through the genotype-to-phenotype map will require a systems-level exploration. The quest for the means by which alternative dimensions of genomic variation are transmitted from one generation to the next and are subject to the forces of natural selection is undoubtedly a promising scientific endeavor.

One of the main reasons why alternative forms of genomic variation have not been adequately explored is the lack of cost-efficient experimental methods to collect data from several individuals in the population. In this perspective, we have pointed out recent developments in long-read sequencing, RNA-seq protocols, and RNA post-transcriptional modifications that will help move the field forward. However, additional methodological advances (e.g., methods allowing the collection of population-level 3D genome structure data in a high-throughput and cost-efficient manner) will be needed to study sources of genomic variation further.

We have here presented a non-exhaustive list of the most promising hidden dimensions of genomic diversity, and believe that all of them need to be taken into account to fully appreciate the evolutionary dynamics and consequences of genomic variation. We encourage the scientific community to remain curious and explore deeper into alternative dimensions of genomic diversity. Altogether, we anticipate that the unraveling of the genotype-to-phenotype map will continue to marvel our minds for years to come.

Acknowledgements

JAR wishes to acknowledge funding from Novo Nordisk Fonden grant NNF-21OC0070726 and HORIZON-TMA-MSCA-PF-EF (POPARCH - PID:101064828). Research by MB-V and CFB is funded by the Swiss National Science Foundation grant 207853. We acknowledge Marc Robinson-Rechavi for his useful comments and feedback. Research by DAH is funded by the Swiss National Science Foundation grant 31003A-163428. Research in AR-H's laboratory is supported by the Spanish Ministry of Economy, Industry and Competitiveness (CGL2014-54317-P, CGL2017-83802-P), the Spanish Ministry of Science and Innovation (PID2020-112557GB-I00 founded by AEI, AEI/10.13039/501100011033); and AGAUR (2021 SGR 0122). MF-M is supported by CONAHCyT grants 303068, 15758 and 137721 and PAPPIT grant IN210323. Research in the LFP lab is funded by the Max Planck Society. JP-T is a member of the International Max Planck Research School 'From Molecules to Organisms'. PM is supported by a fellowship from the EMBL Interdisciplinary Postdoc-Exploring Life in Context (EIPOD-LinC) fellowship program.

Author contributions

MB-V, DAH, and JAR conceptualized the perspective. All authors contributed equally to the writing of the manuscript.

References

Acemel RD, Lupiáñez DG. 2023. Evolution of 3D chromatin organization at different scales. Curr. Opin. Genet. Dev. 78:102019. doi: 10.1016/j.gde.2022.102019.

Albertin CB et al. 2022. Genome and transcriptome mechanisms driving cephalopod evolution. Nat. Commun. 13:2427. doi: 10.1038/s41467-022-29748-w.

Álvarez-González L, Burden F, et al. 2022. 3D chromatin remodelling in the germ line modulates genome evolutionary plasticity. Nat. Commun. 13:2608. doi: 10.1038/s41467-022-30296-6.

Álvarez-González L, Arias-Sardá C, et al. 2022. Principles of 3D chromosome folding and evolutionary genome reshuffling in mammals. Cell Rep. 41:111839. doi: 10.1016/j.celrep.2022.111839.

Álvarez-González L, Ruiz-Herrera A. 2024. Evolution of 3D Chromatin Folding. Annu Rev Anim Biosci. 13. doi: 10.1146/annurev-animal-111523-102233.

Barbosa-Morais NL et al. 2012. The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science. 338:1587–1593. doi: 10.1126/science.1230612.

Borodin P et al. 2022. Mendelian nightmares: the germline-restricted chromosome of songbirds. Chromosome Res. 30:255–272. doi: 10.1007/s10577-022-09688-3.

Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: from polygenic to omnigenic. Cell. 169:1177. doi: 10.1016/j.cell.2017.05.038.

Brasó-Vives M et al. 2022. Parallel evolution of amphioxus and vertebrate small-scale gene duplications. Genome Biol. 23:243. doi: 10.1186/s13059-022-02808-6.

Bucao CF et al. 2024. Inter-individual gene expression variability implies stable regulation of brain-biased genes across organs. bioRxiv. doi: 10.1101/2024.11.11.623020.

Cañestro C, Albalat R, Irimia M, Garcia-Fernàndez J. 2013. Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. Semin. Cell Dev. Biol. 24:83–94. doi: 10.1016/j.semcdb.2012.12.008.

Cano AV et al. 2023. Mutation bias and the predictability of evolution. Philos. Trans. R. Soc. B Biol. Sci. 378:20220055. doi: 10.1098/rstb.2022.0055.

Clarence T et al. 2023. Robust 3D modeling reveals spatiosyntenic properties of animal genomes. iScience. 26:106136. doi: 10.1016/j.isci.2023.106136.

Dixon JR et al. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 485:376–380. doi: 10.1038/nature11082.

Elliott TA, Gregory TR. 2015. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philos. Trans. R. Soc. B Biol. Sci. 370:20140331. doi: 10.1098/rstb.2014.0331.

Eres IE, Gilad Y. 2020. A TAD Skeptic: Is 3D Genome Topology Conserved? Trends Genet. 37:216. doi: 10.1016/j.tig.2020.10.009.

Franke M et al. 2021. CTCF knockout in zebrafish induces alterations in regulatory landscapes and developmental gene expression. Nat. Commun. 12:5415. doi: 10.1038/s41467-021-25604-5.

Harewood L, Fraser P. 2014. The impact of chromosomal rearrangements on regulation of gene expression. Hum. Mol. Genet. 23:R76–R82. doi: 10.1093/hmg/ddu278.

Hartasánchez DA et al. 2023. Expression of cell-wall related genes is highly variable and correlates with sepal morphology. Peer Community J. 3:e93. doi: 10.24072/pcjournal.327.

Hartasánchez DA, Brasó-Vives M, Heredia-Genestar JM, Pybus M, Navarro A. 2018. Effect of collapsed duplications on diversity estimates: what to expect. Genome Biol. Evol. 10:2899–2905. doi: 10.1093/gbe/evy223.

Heger P, Marin B, Bartkuhn M, Schierenberg E, Wiehe T. 2012. The chromatin insulator CTCF and the emergence of metazoan diversity. Proc. Natl. Acad. Sci. 109:17507–17512. doi: 10.1073/pnas.1111941109.

Hoencamp C et al. 2021. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. Science. 372:984–989. doi: 10.1126/science.abe2218.

Irimia M et al. 2012. Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res. 22:2356–2367. doi: 10.1101/gr.139725.112.

Jeffares DC et al. 2015. The genomic and phenotypic diversity of Schizosaccharomyces pombe. Nat. Genet. 47:235–241. doi: 10.1038/ng.3215.

Ji P, Murata-Hori M, Lodish HF. 2011. Formation of mammalian erythrocytes: chromatin condensation and enucleation. Trends Cell Biol. 21:409–415. doi: 10.1016/j.tcb.2011.04.003.

Kaushal A et al. 2021. CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions. Nat. Commun. 12:1011. doi: 10.1038/s41467-021-21366-2.

LaBoone PA, Assis R. 2024. Stress-Induced Constraint on Expression Noise of Essential Genes in E. coli. J. Mol. Evol. doi: 10.1007/s00239-024-10211-x.

Lewin HA et al. 2022. The Earth BioGenome Project 2020: Starting the clock. Proc. Natl. Acad. Sci. 119:e2115635118. doi: 10.1073/pnas.2115635118.

Li M et al. 2022. De Novo Assembly of 20 Chicken Genomes Reveals the Undetectable Phenomenon for Thousands of Core Genes on Microchromosomes and Subtelomeric Regions. Mol. Biol. Evol. 39:msac066. doi: 10.1093/molbev/msac066.

Majic P, Erten EY, Payne JL. 2022. The Adaptive Potential of Nonheritable Somatic Mutations. Am. Nat. 200:755–772. doi: 10.1086/721766.

Majic P, Payne JL. 2023. Developmental Selection and the Perception of Mutation Bias. Mol. Biol. Evol. 40:msad179. doi: 10.1093/molbev/msad179.

Martin NS, Schaper S, Camargo CQ, Louis AA. 2024. Non-Poissonian Bursts in the Arrival of Phenotypic Variation Can Strongly Affect the Dynamics of Adaptation. Mol. Biol. Evol. 41:msae085. doi: 10.1093/molbev/msae085.

Matthews LJ, Turkheimer E. 2022. Three legs of the missing heritability problem. Stud. Hist. Philos. Sci. 93:183–191. doi: 10.1016/j.shpsa.2022.04.004.

Misteli T. 2007. Beyond the Sequence: Cellular Organization of Genome Function. Cell. 128:787–800. doi: 10.1016/j.cell.2007.01.028.

Mitchell KJ, Cheney N. 2024. The Genomic Code: The genome instantiates a generative model of the organism. arXiv:2407.15908. doi: 10.48550/arXiv.2407.15908.

Mitschka S, Mayr C. 2022. Context-specific regulation and function of mRNA alternative polyadenylation. Nat. Rev. Mol. Cell Biol. 23:779–796. doi: 10.1038/s41580-022-00507-5.

Mota-Gómez I et al. 2022. Sex-determining 3D regulatory hubs revealed by genome spatial auto-correlation analysis. bioRxiv. doi: 10.1101/2022.11.18.516861.

Murphy ZC et al. 2021. Regulation of RNA polymerase II activity is essential for terminal erythroid maturation. Blood. 138:1740–1756. doi: 10.1182/blood.2020009903.

Necsulea A et al. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature. 505:635–640. doi: 10.1038/nature12943.

Niu L et al. 2021. Three-dimensional folding dynamics of the Xenopus tropicalis genome. Nat. Genet. 53:1075–1087. doi: 10.1038/s41588-021-00878-z.

Nora EP et al. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell. 169:930-944.e22. doi: 10.1016/j.cell.2017.05.004.

Pallares LF et al. 2023. Saturating the eQTL map in Drosophila melanogaster: genome-wide patterns of cis and trans regulation of transcriptional variation in outbred populations. bioRxiv. doi: 10.1101/2023.05.20.541576.

Pallares LF. 2019. Searching for solutions to the missing heritability problem. eLife. 8:e53018. doi: 10.7554/eLife.53018.

Pallares LF, Picard S, Ayroles JF. 2020. TM3'seq: A Tagmentation-Mediated 3' Sequencing Approach for Improving Scalability of RNAseq Experiments. G3 GenesGenomesGenetics. 10:143–150. doi: 10.1534/g3.119.400821.

Pauli T et al. 2016. Transcriptomic data from panarthropods shed new light on the evolution of insulator binding proteins in insects. BMC Genomics. 17:861. doi: 10.1186/s12864-016-3205-1.

Paulsen T, Kumar P, Koseoglu MM, Dutta A. 2018. Discoveries of Extrachromosomal Circles of DNA in Normal and Tumor Cells. Trends Genet. 34:270–278. doi: 10.1016/j.tig.2017.12.010.

Penagos-Puig A et al. 2023. RNA polymerase II pausing regulates chromatin organization in erythrocytes. Nat. Struct. Mol. Biol. 30:1092–1104. doi: 10.1038/s41594-023-01037-0.

Rao SSP et al. 2017. Cohesin Loss Eliminates All Loop Domains. Cell. 171:305-320.e24. doi: 10.1016/j.cell.2017.09.026.

Reyes A, Huber W. 2018. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. Nucleic Acids Res. 46:582–592. doi: 10.1093/nar/gkx1165.

Rogers TF, Simakov O. 2023. Emerging questions on the mechanisms and dynamics of 3D genome evolution in spiralians. Brief. Funct. Genomics. 22:533–542. doi: 10.1093/bfgp/elad043.

Rowley MJ et al. 2017. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol. Cell. 67:837-852.e7. doi: 10.1016/j.molcel.2017.07.022.

Rozhoňová H, Martí-Gómez C, McCandlish DM, Payne JL. 2024. Robust genetic codes enhance protein evolvability. PLOS Biol. 22:e3002594. doi: 10.1371/journal.pbio.3002594.

Saitou M, Gokcumen O. 2020. An Evolutionary Perspective on the Impact of Genomic Copy Number Variation on Human Health. J. Mol. Evol. 88:104–119. doi: 10.1007/s00239-019-09911-6.

Santos-Rodriguez G, Voineagu I, Weatheritt RJ. 2021. Evolutionary dynamics of circular RNAs in primates. eLife. 10:e69148. doi: 10.7554/eLife.69148.

Sigalova OM, Shaeiri A, Forneris M, Furlong EE, Zaugg JB. 2020. Predictive features of gene

expression variation reveal mechanistic link with differential expression. Mol. Syst. Biol. 16:e9539. doi: 10.15252/msb.20209539.

Sundaram V, Wysocka J. 2020. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. Philos. Trans. R. Soc. B Biol. Sci. 375:20190347. doi: 10.1098/rstb.2019.0347.

Tautz D, Reeves G, Pallares LF. 2020. New experimental support for long standing concepts of polygenic genetics implies that the Mendelian genetic paradigm needs to be revised. NAL-Live. 2020:1–15. doi: 10.34714/leopoldina_nal-live_0001_01000.

Tsour S et al. 2024. Alternate RNA decoding results in stable and abundant proteins in mammals. bioRxiv. doi: 10.1101/2024.08.26.609665.

Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. Nat. Rev. Genet. 18:411–424. doi: 10.1038/nrg.2017.26.

Vara C et al. 2021. The impact of chromosomal fusions on 3D genome folding and recombination in the germ line. Nat. Commun. 12:2981. doi: 10.1038/s41467-021-23270-1.

Vara C, Ruiz-Herrera A. 2022. Unpacking chromatin remodelling in germ cells: implications for development and evolution. Trends Genet. 38:422–425. doi: 10.1016/j.tig.2021.10.007.

Vollger MR et al. 2022. Segmental duplications and their variation in a complete human genome. Science. 376:eabj6965. doi: 10.1126/science.abj6965.

Wagner A. 2012. The role of robustness in phenotypic adaptation and innovation. Proc. R. Soc. B Biol. Sci. 279:1249–1258. doi: 10.1098/rspb.2011.2293.

Whitehead DJ, Wilke CO, Vernazobres D, Bornberg-Bauer E. 2008. The look-ahead effect of phenotypic mutations. Biol. Direct. 3:18. doi: 10.1186/1745-6150-3-18.

Wolf S et al. 2023. Characterizing the landscape of gene expression variance in humans. PLOS Genet. 19:e1010833. doi: 10.1371/journal.pgen.1010833.

Wright CJ, Smith CWJ, Jiggins CD. 2022. Alternative splicing as a source of phenotypic diversity. Nat. Rev. Genet. 23:697–710. doi: 10.1038/s41576-022-00514-4.

Yanagida H et al. 2015. The Evolutionary Potential of Phenotypic Mutations. PLOS Genet. 11:e1005445. doi: 10.1371/journal.pgen.1005445.

Yu H et al. 2024. Pan-evolutionary and regulatory genome architecture delineated by an integrated macro- and microsynteny approach. Nat. Protoc. 19:1623–1678. doi: 10.1038/s41596-024-00966-4.