

REVIEW

Open Access



Placing human gene families into their evolutionary context

Alex Dornburg^{1*}, Rittika Mallik¹, Zheng Wang², Moisés A. Bernal³, Brian Thompson⁴, Elspeth A. Bruford^{5,6}, Daniel W. Nebert^{7,8}, Vasilis Vasiliou⁴, Laurel R. Yohe¹, Jeffrey A. Yoder⁹ and Jeffrey P. Townsend^{1,10}

Abstract

Following the draft sequence of the first human genome over 20 years ago, we have achieved unprecedented insights into the rules governing its evolution, often with direct translational relevance to specific diseases. However, staggering sequence complexity has also challenged the development of a more comprehensive understanding of human genome biology. In this context, interspecific genomic studies between humans and other animals have played a critical role in our efforts to decode human gene families. In this review, we focus on how the rapid surge of genome sequencing of both model and non-model organisms now provides a broader comparative framework poised to empower novel discoveries. We begin with a general overview of how comparative approaches are essential for understanding gene family evolution in the human genome, followed by a discussion of analyses of gene expression. We show how homology can provide insights into the genes and gene families associated with immune response, cancer biology, vision, chemosensation, and metabolism, by revealing similarity in processes among distant species. We then explain methodological tools that provide critical advances and show the limitations of common approaches. We conclude with a discussion of how these investigations position us to gain fundamental insights into the evolution of gene families among living organisms in general. We hope that our review catalyzes additional excitement and research on the emerging field of comparative genomics, while aiding the placement of the human genome into its existentially evolutionary context.

Introduction

Human genomes contain groups of genes that share common ancestry and are often functionally related. These *gene families* comprise the largest proportion of the protein-coding sequences, and in many cases, they are the product of gene duplications over time. However, characterizing the sequence and functional diversity of gene families among species has been challenging. Organisms can show sequence divergence among duplicate genes, as well as differences in numbers of genes within each gene family; these variations can be observed among closely

related species or even at the population level (i.e., copy number variation). This molecular diversity is the manifestation of a complex evolutionary history of gene duplication over time that can result in two or more paralogs on the same chromosome (i.e., *cis*) that are either clustered or at distant chromosomal locations. Alternatively, paralogs can occur on different chromosomes (i.e., *trans*) or reflect a combination of these locations. Because many of these duplications date to deep evolutionary divergences, efforts to decode human gene families rely on interspecific comparative genomic studies between humans and other animals [1–4].

Early research comparing the genomes of model species (i.e., fruit flies, mice, etc.) to humans has yielded tremendous benefits to our understanding of human genomics. This process has been markedly enhanced by the recent proliferation of genomic sequence, annotation, and

*Correspondence: adornbur@uncc.edu

¹ Department of Bioinformatics and Genomics, UNC-Charlotte, Charlotte, NC, USA

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

transcriptomic resources for non-model species [5]. This renaissance of comparative approaches is rapidly linking human gene families and their respective functions, to a wide variety of species that previously lacked genomic resources. This expansion on the genetic resources available for non-model species has also revealed gene families that are absent in humans, but that have analogous functions in the latter. By understanding how these gene families evolve, we can reveal the processes, synergies, and constraints governing genome evolution across the Tree of Life, while providing a better understanding of the gene families in the human genome. In conjunction, these insights are crucial for informing and empowering experimental research in years to come.

In this review, we reflect on the growing body of knowledge associated with gene family evolution. We begin by highlighting the parallels between comparative genomics and phylogenomics research. We then review cases in which comparative genomic approaches have advanced our understanding of how gene families contribute to the evolution of gene expression, as well as the roles that different gene families play in the interactions of species with their biotic and abiotic environments. Next, we survey recent methodological advances and challenges and conclude by highlighting emergent research frontiers and the limits that persist for this type of research. Comparative genomic studies will continue to grow in number in parallel with the growth of high-quality genomic resources. We argue that we will be able to not only characterize gene family diversity but also decipher meaningful functional implications from variation in gene duplication, thereby quantitatively measuring the specific processes that underlie gene family evolution.

Placing the human genome into context

The current reference assembly of the human genome, GRCh38, constitutes ~3100 megabases that are predicted to encode up to 20,000 protein-coding genes, over 25,000 non-protein-coding genes, and an astonishing ~15,000 pseudogenes (i.e., non-functional copies of protein-coding genes). Around three-quarters of human protein-coding genes have at least one recognizable paralog, i.e., a homologous gene that arose following duplication of an ancestral gene. This high frequency of paralogous genes indicates that much of the coding human genome can be grouped into gene families (i.e., sets of homologous genes). Because of the deep ancestry of gene duplication and divergence, many of the gene families found in humans are either homologous to, or analogous with, the diversity of genes found across the Tree of Life. Consequently, the biology of human gene families can be illuminated both by their study in humans and by their comparison with other organisms. Phylogenetics

provides the inferential machinery to illuminate evolution across these macro- and microscales, by informing the relation of species trees to the organismal evolution of species and the relation of gene trees to the molecular evolution of genes. Correspondingly, application of phylogenetics to our evolutionary history has shed light on the origins of gene families within *Homo sapiens*.

Phylogenetic approaches can play an important role in revealing the processes that underlie the origins and function of human gene families. Intuitively, gene family evolution is expected to reflect the evolutionary relationships among species, with divergences arising as a result of selective pressures, mutations and drift across hundreds of millions of years of evolution [6]. However, this gradual and continuous evolutionary change is not always manifest. Some gene families are more labile and can actively generate novel sequence and functional diversity, while others are highly constrained. The latter represent crucial components of the genetic architecture within the Tree of Life that constitute a historical archive of sequence and functional diversity. Though representing a conceptually useful dichotomy, these two extremes are not mutually exclusive. We can expect some gene families with tightly conserved functions to also contribute novel paralog templates that facilitate sequence diversity and innovation. Just as the inference of species trees enables us to understand functional parallels and divergences at the organismal level (i.e., macroscale), inference of gene family trees across multiple species enables us to understand functional parallels and divergences at the gene level (i.e., microscale). The rapid accumulation of genome sequences for lineages across the Tree of Life now empowers such inferences.

The evolution of gene expression

Some of the first insights into the evolution of gene expression came from the study of gene families. Expression of gene duplicates has been shown to evolve at as much as a tenfold higher rate immediately following duplication, thereafter slowing, and to do so asymmetrically and leading to increasing gene network complexity [7]. The evolution of gene expression is known to be associated with complex gene regulatory networks (GRNs), and divergence in the former often produces changes in the latter [8]. Overall, changes in GRNs and their accompanying differences in expression play an essential role in the phenotypic diversity observed throughout the Tree of Life, making these studies highly relevant for genetics, physiology, morphology, and even ecology. Despite this clear relevance for the observed diversity, there are many challenges for directly comparing gene expression among multiple species. First, differential gene expression is expected to change significantly with genetic distance

[9, 10]. Even when the relationship is not directly proportional, studies have shown that differential gene expression between even distant populations of the same species can be elevated even under “Control” experimental conditions [9]. Second, the expression dosage level of a duplicated gene is expected to be twofold; however, functional genomics experiments have demonstrated that gene duplicates can be expressed at over 2–5 times the levels of single-copy genes [11]. Third, there are many logistical issues for conducting common-garden experiments with multiple species. In addition to ethical considerations concerning human and non-human research subjects, the need to account for tissue homologies (e.g., quantitative differences in tissue content of organs of divergent species) is another challenge that arises. This challenge renders the practice of comparing expression of related genes within a single species easier than comparisons of the expression of related genes among divergent species. Newly emergent computational tools coupled with phylogenetic approaches are now allowing us to overcome such challenges. For example, a phylogenetic analysis of expression variance and evolution (EVE) is able to provide information on differences in gene expression promoted by plasticity relative to those associated with genetic divergence [12]. We are now positioned to test fundamental hypotheses concerning the evolution of gene expression [13].

Studies have suggested that gene expression evolves mostly through stabilizing selection [14], with an apparent paucity of directional selection. Estimates indicate expression divergence is usually below expectations from random mutations and genetic drift between even highly divergent species such as fruit flies, mice, and primates [15], a result often attributed to compensatory effects between *cis*- and *trans*-regulatory elements of expression [16]. Correspondingly, investigations have revealed lower differences in expression than expected between closely related species [17]. In contrast, uncommon cases of directional selection caused by mutations in coding sequences, changes in gene regulatory regions, and gene duplications are responsible for large effects on phenotype and function among divergent species [8, 18, 19]. Continued comparative investigations on the evolution of gene expression represents a fruitful avenue that could illuminate mechanisms underlying rapid phenotypic diversification.

The impact of environmental change on genomic evolution

All organisms respond in some measure to changes in external abiotic conditions. Due to the fast pace of changes promoted by anthropogenic activities, the study of how abiotic fluctuations impact gene expression has increased

dramatically in the past decades. Thus, comparative studies of gene expression between species have emerged as a critical basis for understanding how organisms respond to environmental changes. Despite the intrinsic differences between metabolic regulators and conformers, studies have found overlap in the functions that can be activated in the face of environmental stress. In particular, a great deal of overlap has been detected in mechanisms associated with metabolic compensation, as well as responses to cellular stress. These insights now position us to better predict potential genomic consequences to specific environmental changes.

How lineages respond to temperature increase has become an area of increased research focus, where there is overlap of implications between humans and other species. In the particular case of aquatic poikilotherms, increasing temperatures lead to an increase in the metabolic activity, which in turn leads to an elevated demand for oxygen consumption [20, 21]. This increase in oxygen consumption as a function of aerobic demand can lead to enhanced cellular stress due to the elevation of reactive oxygen species, which can trigger a cascade of detrimental effects for cells. In the liver, for example, responses of fishes to elevated temperature have been associated with the activation of gene categories related to the mTOR complex, oxidation–reduction, cellular apoptosis, mitochondrial electron transport chain, permeability of the endoplasmic reticulum, etc., which are also extensively reported to be activated in humans and mice exposed to intense exercise [22–24]. Similar observations have been reported in fish brains exposed to elevated temperatures, where activation of compensatory mechanisms can be similar to those observed in humans with cerebrovascular strokes or other neuro-motor ailments. For example, *Drosophila* flies and fish exposed to warm environments for multiple generations can exhibit compensation by improving neuro-motor connections by the activation of the gene plastin 3, which in humans and mice is a protective modifier for spinal muscular atrophy [25, 26]. Similarly, warm water conditions can lead to the activation of genes associated with low oxygen conditions in the nervous system of aquatic animals, as a result of the increased oxygen demand, and similar pathways are known to be activated in humans with strokes and circulatory deficiencies (e.g., neuroglobins) [26]. Given climate forecasts for the next century, further investigation of how gene families evolve under changing abiotic conditions will be vital for identifying potential future risks to public health.

Immune gene families martial a clustered frontline of defense against disease

Many innate immune receptors involved in differentiating between self and non-self are encoded by gene families that are organized as clusters throughout the genome

and encode both activating and inhibitory forms. For such families of clustered genes, high sequence diversity and even diversity in the presence or absence of gene orthologs between humans and more distantly related vertebrates should be expected. Such a fast pace of molecular evolution is necessary to compete in the evolutionary arms race with rapidly evolving pathogenic threats [27]. However, there is also significant diversity in sequences and even genes between individual humans. For instance, the leukocyte receptor complex (LRC) is likely one of the best studied examples of intraspecific gene content variation, in which different people encode different numbers and combinations of killer cell immunoglobulin-like receptor (KIR) genes which play important roles in natural killer (NK) cell function [28] (Fig. 1A). A large-scale study that defined the KIR haplotypes for 800 families of European origin resolved more than 3000 haplotypes and identified over 70 unique haplotypes based solely on gene content [29]. Different KIR gene content haplotypes have been linked to differences in susceptibility to infectious diseases such as HIV and malaria [30–32] as well as autoimmune disorders [33] and are important factors for the success of hematopoietic stem cell transplantation for leukemia patients [34].

Clustered families of innate immune receptors are not unique to humans. However, expansions and contractions of these gene families vary substantially across the Tree of Life. For example, mice do not have an expanded cluster of KIRs. Instead, gene family variation similar to human KIRs is evident in the natural killer complex (NKC) of mice, in which different inbred mouse strains encode different numbers and combinations of Ly49 genes, also known as killer cell lectin-like receptors or Klras. In rodents, these genes play similar roles in NK cells as KIRs do in humans [35] (Fig. 2A). The diversity of Ly49/Klra gene cluster content in mice stands in contrast to the single *KLRA1P* pseudogene in the human genome and is linked to differences in susceptibility to murine cytomegalovirus (MCMV) infection [36]. Given the intraspecific gene content variation that is observed for numerous gene families in humans (KIRs) and mice (Ly49/Klra), this discrepancy challenges us to consider which vertebrate lineages possess gene family expansions convergent with, or functionally analogous to human expansions. Moreover, a comparative perspective enables us to derive general rules that underlie the diversification of clustered vertebrate immune genes and the maintenance of specific levels of intraspecific variation.

In the case of the LRC, KIRs display dramatic convergences in gene content variation between species. Humans encode 4–20 KIR genes (including rare haplotypes [29]) and cows encode up to 18 KIR genes, whereas multiple other vertebrates encode one or two KIRs, or,

like mice, have lost all KIR genes at the LRC (Table 1) [1, 37–40] (Fig. 1B). However, studies on the *intraspecific* gene content variation of the LRC are limited primarily to KIRs in primates and very little is known about this level of structural variation in other mammalian lineages. Further, the LRC contains multiple gene families that encode innate immune receptors structurally similar to KIRs, such as the leukocyte Ig-like receptors (LILRs) and leukocyte-associated Ig-like receptors (LAIRs) [41, 42]. These LRC loci have been associated with multiple immunological disorders including rheumatoid arthritis, multiple sclerosis, and lupus [43], are expressed on a range of immune cell types, and have been classified as activating or inhibitory based. Moreover, the LILR genes in the LRC also display interspecific gene content variation. Respectively, reference genomes for humans, cows, horses, and elephants encode 11, 26, 3, and 2 orthologs of LILR genes [38, 39, 44] with mice also encoding at least 8 LILR orthologs (also referred to as paired Ig-like receptors or PIRs) [45, 46]. Even among LAIRs—for which humans encode only two genes (*LAIR1* and *LAIR2*)—three LAIR-like sequences have been reported in pig [47] and elephant [38], and Ensembl predicts that a number of mammalian species encode paralogs of *LAIR1* (e.g., black snub-nosed monkey, vervet, drill, greater horseshoe bat) or have lost *LAIR1* (e.g., polar bear, armadillo, shrew, dolphin).

Expansions of clustered gene families such as these LRC loci are predicted to increase receptor diversity, an increase that can enable a better defense against the next unknown pathogen. Why these lineage-specific expansions are heterogeneously distributed across the vertebrate Tree of Life remains less clear. A potential explanation may lie in functional overlap. For example, KIR and Ly49/Klra proteins are considered functional analogs, as both interact with MHC class I proteins, mediating the recognition and direct killing of infected and cancerous cells. The NKC gene cluster, which includes the Ly49/Klra genes, displays dramatic interspecific gene-content diversity (Fig. 2B). There is a single Ly49/Klra locus in humans and dogs, compared to 6 and up to 21 Ly49/Klra genes in horses and mice, respectively (Table 2) [39, 48, 49].

As divergences between humans and other vertebrates span deeper timescales, finding functional analogs becomes increasingly challenging. Initial studies looking to identify mouse KIRs led to conflicting results [50, 51]. We now know that KIRs and Ly49/Klras are structurally different proteins and do not share a common genetic origin, but play comparable roles in NK cells' surveillance for transformed and infected cells. As mentioned above, humans encode a single Ly49/Klra pseudogene (*KLRA1P*) and mice have lost the KIR genes from the LRC. This

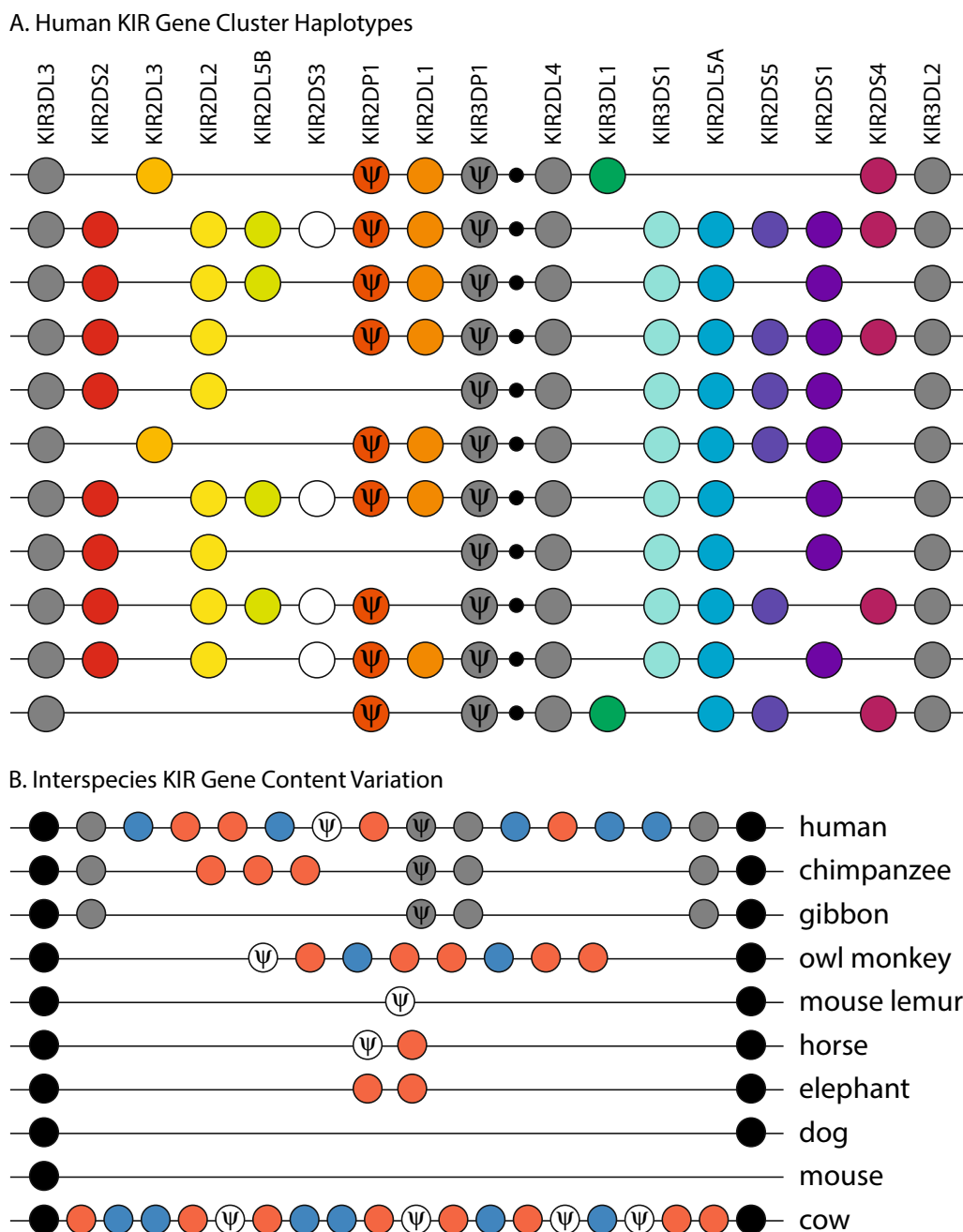
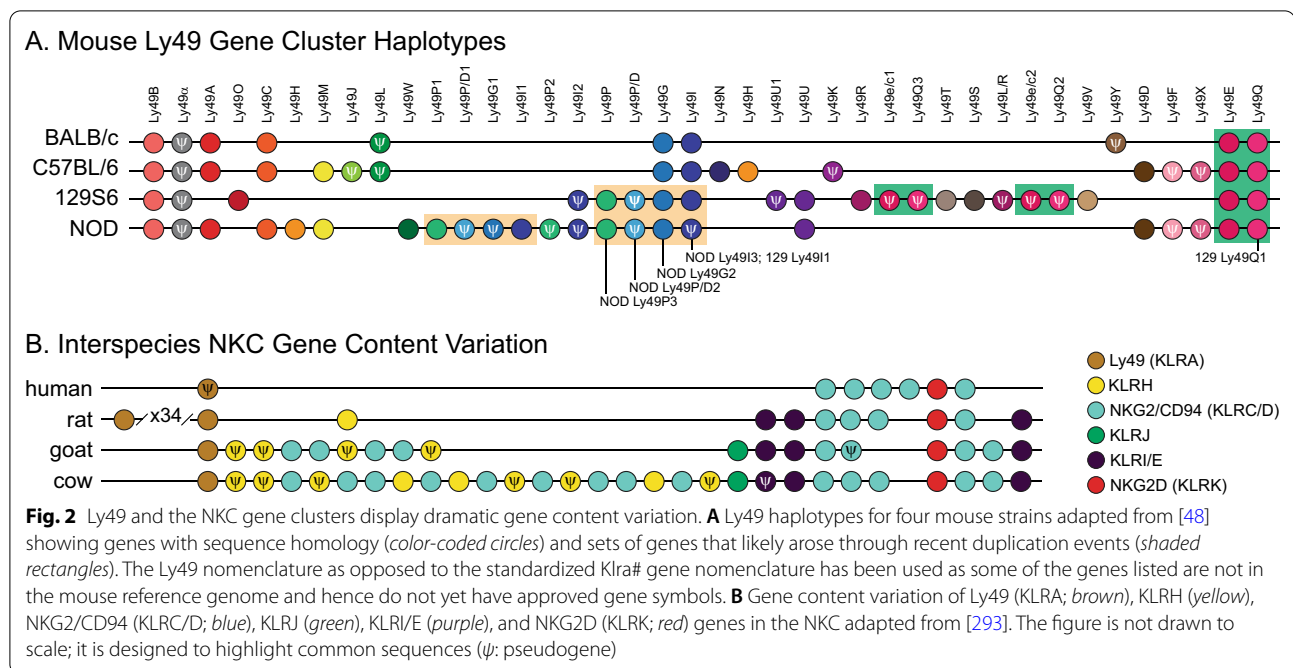


Fig. 1 Killer cell Ig-like receptor gene clusters display dramatic gene content variation. **A** Eleven gene content haplotypes for the human killer cell Ig-like receptor cluster within the leukocyte receptor complex adapted from Middleton and Gonzelez [28]. Framework killer cell Ig-like receptor genes are conserved across haplotypes (gray circles), whereas other genes (color-coded circles) are variably present across haplotypes. Additional haplotypic variation is achieved through a recombination hotspot between *KIR3DP1* and *KIR2DL4* [29] (small black circle). **B** Variation in the number and combination of killer cell Ig-like receptor genes within the leukocyte receptor complex in mammalian genomes [37–40]. Framework killer cell Ig-like receptor genes are conserved in primates (gray circles) and bounded by conserved flanking genes (black circles). The numbers of killer cell Ig-like receptors with predicted inhibitory (red) and activating (blue) functions vary between species. Additional gene content variation is observed for other gene families within the leukocyte receptor complex (e.g., leukocyte Ig-like receptors and leukocyte-associated Ig-like receptors). The figure is not to drawn scale; it is designed to highlight common sequences (ψ : pseudogene)



extreme case of convergent evolution highlights what can be missed when only two species are compared. Initial hypotheses, prior to the completion of the sequencing of the human genome, was that primates used an expanded set of KIR genes and rodents used an expanded set of Ly49/Klra genes for the same NK function and that an expansion of the KIR or Ly49/Klra genes was required for a functionally competent immune system. As genomic sequencing became more affordable and the LRC and NKC were sequenced from a wider range of mammals, it became obvious that certain lineages expanded the KIR cluster (e.g., primates and cattle), while others expanded the Ly49/Klra cluster (rodents and equidae), and yet others encode a single KIR and a single Ly49/Klra gene (e.g., pinnipeds) [49]. This observation in seals and sea lions made it clear that the long-term survival of placental mammals does not require an expanded system of either Ly49/Klra or KIR receptors.

Natural killer (NK) cells have been described as one of the oldest immune cell types that differentiates between self and non-self. Although NK cells have been functionally described from bony fish [52, 53], numerous efforts to identify genes encoding KIRs or Ly49/Klra in fish have been unsuccessful [1]. Nevertheless, in the same way that humans use KIRs and mice use Ly49/Klras to mediate NK function, a candidate gene family that may facilitate mammalian-like NK cell function in fishes is the novel immune-type receptors (NITRs) gene family [54–57]. Similar to mammalian KIRs, NITRs are encoded in gene clusters, include inhibitory and activating forms,

and display gene content variation [1, 55, 58]. NITRs have recently been found to have ancient evolutionary origins within the earliest divergences of ray-finned fishes, an evolutionary persistence that lends support to the hypothesis that this receptor family offers some core immune response functionality [59–62]. Unfortunately, it is not known if NITR evolution is analogous to that of KIRs and Ly49/Klras. This lack of understanding imposes limits on our ability to link NK cell function in teleost model organisms (e.g., zebrafish, stickleback, and medaka) to the mammalian immune system. Because there are numerous immune receptor families that display both inter- and intraspecific gene-content variation across ray-finned fishes and are not identifiable in mammalian (or other tetrapod) genomes [59, 60, 63–65], comparative genomic investigations of the molecular basis of self-recognition pathways in “the other half” of all living vertebrates are an exceptionally rich research frontier that can aid us in understanding the evolution of our own genomes.

Molecular biology and its translation to action against cancer

Genomic analyses of human cancers have revealed several phenotypes that are shared across a vast majority of types and have been dubbed “cancer hallmarks” [66, 67]. These hallmarks include immune evasion, resisting cell death, and evasion of growth suppressors to name but a few. In many cases, the genetic basis of these and other cancer hallmarks can be linked to gene families

that are broadly distributed among mammals or other vertebrates. For example, the TP53 (p53) gene family arguably represents the most well-studied family of human oncogenes [68]. This gene family has evolutionary origins dating back to the most recent common ancestor of choanoflagellates and metazoans, where a single gene involved in germline DNA repair has been maintained [69]. This gene duplicated multiple times during the evolution of jawed vertebrates, giving rise to *TP53*, *TP63*, and *TP73*. While decades of comparative research have defined general features of this gene family's evolution and biology [68], recent comparative studies have illuminated the role of this gene family in mitigating the risk of cancer development in the largest land mammals—elephants.

Across vertebrates, there is a positive relationship between the risk of developing cancers and increases in body size [70]. This relationship is a consequence of a larger cell count increasing the chance of malignant cell transformations [71]. Additionally, the risk of developing cancer is positively correlated with longevity [72]. As a consequence of their large-body size and long life spans, elephants should be expected to have some of the highest rates of cancer among terrestrial mammals. However, this expectation is contradicted by empirical data. Elephants achieve a seemingly paradoxical low rate of cancer incidence in part through a duplication of TP53 genes that provide a greater sensitivity to DNA damage [68]. Recent studies have shed light on genes interacting with elephant TP53, revealing that the *LIF* (leukemia inhibitory factor) gene has undergone segmental duplication in selected species including the African elephant. Across analyzed genomes, most copies of this gene are pseudogenized. Elephants stand out as an exception, with investigators finding one gene (*LIF6*) that had refunctionalized with apoptotic function [73]. These studies and others like them have created an exciting opportunity to investigate the genetic basis of DNA damage repair mechanisms.

Beyond the genetics of DNA repair mechanisms, it is also clear that studies of other oncogenes between species can inform the development of new therapies. To effectively guide research in cancer biology, it is important to place these oncogenes into a phylogenetic context. For example, the ALK receptor tyrosine kinase gene (previously “anaplastic lymphoma kinase”) (*ALK*), a prominent receptor tyrosine kinase (RTK) proto-oncogene, is present across a range of metazoans that spans humans to fruit flies with a high degree of structural conservation [74, 75]. *ALK* has been associated with tumorigenesis in neuroblastoma, non-small cell lung cancer, anaplastic large-cell lymphoma, breast and renal cell carcinomas and identified as a potential target for therapeutic development [76–78]. Although studies of model

organisms have revealed fundamental insights into the biology of oncogenic alterations, linking these heterogeneous insights to human cancers was challenged by a lack of a comparative framework. It has recently been demonstrated that in the early history of jawed vertebrates, an ancestral *ALK* duplicated to give rise to the leukocyte receptor tyrosine kinase (*LTK*) gene, and that subsequently *ALK* and *LTK* have traded functional roles between ray-finned fishes and sarcopterygians (e.g., tetrapods, lungfish, and coelacanth). Additionally, even as the homology among *ALK*-like genes becomes increasingly clear, the ligands for ALK in non-vertebrates {Jelly belly (*Jeb*) in *Drosophila melanogaster* [79], hesitation behavior-1 (*hen-1*) in *Caenorhabditis elegans* [80], and *ALKALI/2* (augmentor/FAM150) in humans [81, 82]} are not similarly homologous [74]. This example highlights the critical need for an evolutionary perspective in model organism-based oncogene research to effectively translate findings regarding homologous genes and their interacting partners.

Studies of oncogenes have been crucial in revealing how cancer co-opts existing gene families to enable its persistence. For instance, studies of the genetic mechanisms of tumor growth or DNA repair remain invaluable. However, an additional frontier in cancer biology leverages comparative immunogenetics to investigate the evolution of receptors that modulate the adaptive and immune system response, to better understand how cancers evade detection or elimination. An example of how cancer co-opts the genetic machinery of the mammalian immune system has been revealed by studies of signal regulatory proteins (SIRPs), a family of transmembrane glycoproteins with extracellular immunoglobulin-like domains, involved in the regulation of tyrosine kinase-coupled signaling processes [83, 84]. SIRPs have been identified broadly among mammals including humans, primates, rodents, dogs, cows, horses, and opossum [85, 86]. In humans, the SIRP family contains the inhibitory SIRP α , activating SIRP β , non-signaling SIRP γ , and soluble SIRP δ encoded in a gene cluster on chromosome 20 [83, 86, 87]. Putative clusters with different numbers of SIRP homologs have been observed in other mammals, birds (*Gallus*), and lizards (*Anolis*) [85, 86, 88]. The evolutionary origins of SIRPs remain unknown, as does the scope of predicted functional conservation.

Multiple cancer cells utilize the relationship between SIRP α and its ligand CD47 to prevent tumor cell phagocytosis [89, 90]. CD47 is commonly overexpressed on the cell surface of many cancers, providing a “Don't eat me” signal that engages SIRP α on macrophages and prevents phagocytosis [91, 92]. SIRP α recognizes the elevated levels of CD47 on tumor cells and negatively controls effector function which prevents destruction of the cancer

cells [93, 94]. However, experimental administration of monoclonal antibodies or soluble SIRP α (Fc-fusion) that bind CD47 and block the CD47-SIRP α signaling pathway promotes tumor cell phagocytosis, inhibits tumor growth in mice, and increases survival [92, 95]. A large-scale comparative perspective of immune gene families such as SIRPs and their interacting partners is therefore of high value for the design of new therapies and for revealing natural systems with high translational relevance.

Comparative genomics illuminates the ancestry and diversity of vision

Light sensing is one of the most ancient characteristics of life [96], and proteins associated with human vision have been studied for over a century in various model organisms [97, 98]. These loci have been used in studies that span investigations of molecular evolutionary rates [99], the reconstruction of ancestral proteins in extinct species [100], dim-light vision [101, 102], and extensive applications for molecular systematics [103–105]. However, it has only been in the last few decades that large comparative studies have begun placing gene families in the human genome into evolutionary context. The results of these studies have revealed numerous associations between gene family evolution and the ecology of a lineage, as well as a remarkable scope of genotypic diversity across the Tree of Life associated with light sensing phenotypes.

Families of photoreceptors are widely distributed across multiple phyla—including plants and fungi—that rely on those sensory pathways to regulate their responses to environmental changes [106, 107]. Although these lineages do not possess any structures analogous to human eyes, their diversity of light sensor molecules is extraordinary. Comparatively, fungi have a greater diversity of light sensor molecules than humans, each providing sensation of different ranges of ambient light. In addition to retinal binding rhodopsin (like humans) for sensing green light, fungi also feature phytochrome- and flavin-based photoreceptor gene families, providing red- and blue-light sensation [108]. Gene duplications of fungal photoreceptors are present across many species [108], and their expansion has been clearly linked to fungal ecology [106, 109, 110]. For example, duplication and divergence in fungal rhodopsins and opsin-like proteins have been characteristic of clades of pathogenic and non-pathogenic fungal species [111], whereas extreme expansion of the phytochrome gene family has been found in aquatic fungi with multiple paralogous copies [112].

Among gene families associated with human vision, crystallins represent perhaps one of the most iconic examples of an evolutionary bloom with hypothesized lineage-specific gains and losses that are associated

with the ecology of the vertebrate lens [113–115]. Numerous reviews and in-depth studies of human crystallin gene families have been conducted due to their essential role in the maintenance of lens clarity, and variation in human crystallin genes has been linked to cataracts and to vision loss [113, 116–118]. However, the evolutionary history of γ -crystallin genes has only recently come to light and this gene family represents an intriguing empirical example of a gene family considered “on its way to extinction” in humans [119]. A diverse suite of γ -crystallins associated with the generation of a high refractive index are found in species that possess a hard lens with low water content such as fishes [113, 119]. In contrast, convergent losses or loss of function of γ -crystallins have been reported in lineages such as primates or birds that are characterized by soft lenses with high water content [113, 119]. Humans are estimated to have lost function in two-thirds of the γ -crystallins found in some rodents [120]. Therefore, crystallin gene families are particularly well positioned for future studies investigating the evolution of gene loss.

In addition to crystallins, opsins represent another spectacular example of an evolutionary bloom. Since the sequencing of the first opsin gene nearly 40 years ago in bovids [121, 122], over 1000 opsins have been identified in many species including human, fly, mouse, and zebrafish [123]. Detailed examination of these genes has revealed the basis of color vision across the Tree of Life, often demonstrating striking cases of functional convergences [124, 125]. For example, there are correlations between the opsin repertoire and ecological factors [126, 127], gene losses or loss of function associated with trade-offs between sensory systems [128], amino acid convergence among distantly related taxa (i.e., fish; [129–131]), and convergences in the genetic basis of many human retinal degenerative diseases and vision disorders [132, 133]. The diversity of opsin genes is perhaps not surprising when considering the scale of evolutionary divergences that span lineages such as dragonflies, ray-finned fishes, and humans [134, 135]. However, that this diversity of opsin genes across metazoans is expressed not only in the eye but also in the skin and peripheral tissues has only recently become appreciated [136, 137]. Research in the past decade has revealed a diversity of opsin genes expressed in the skin and other organs of animals including humans [136, 138] and made clear that opsin genes carry out functions that expand beyond light reception. An increased focus on the comparative genomics of opsin gene families already has implications for wound healing, hair growth, optogenetics, and metabolic physiology to name but a few [136, 139]. Future comparative genomic investigations are clearly poised

to inform functional experiments that can illuminate the full diversity of these gene families.

Sniffing out the genomic basis of diverse chemosensation receptors

Chemosensation is ubiquitous throughout the animal Tree of Life. However, the molecular basis of chemosensation has only recently been elucidated. Discovery of chemosensation receptor genes occurred within the last two to three decades, and its discoverers were awarded a Nobel Prize in 2004 [140, 141]. The mechanism of chemosensation involves the conversion of environmental chemical information into neurophysiological information in the brain and is relatively similar across vertebrates and invertebrates [142, 143]. Volatile and nonvolatile chemicals interact with proteins encoded by genes of large multi-copy gene families expressed in sensory neurons and supporting cells of the olfactory and gustatory systems (*cf* [143] for review). Molecular investigations of odorant receptors have revealed that these account for the largest gene families in animals, encoding for up to 5% of the protein-coding genome in mammals alone [144–146]. As these gene families are among the fastest evolving in the genome [147], this makes them ideal models for understanding complex patterns of gene family evolution.

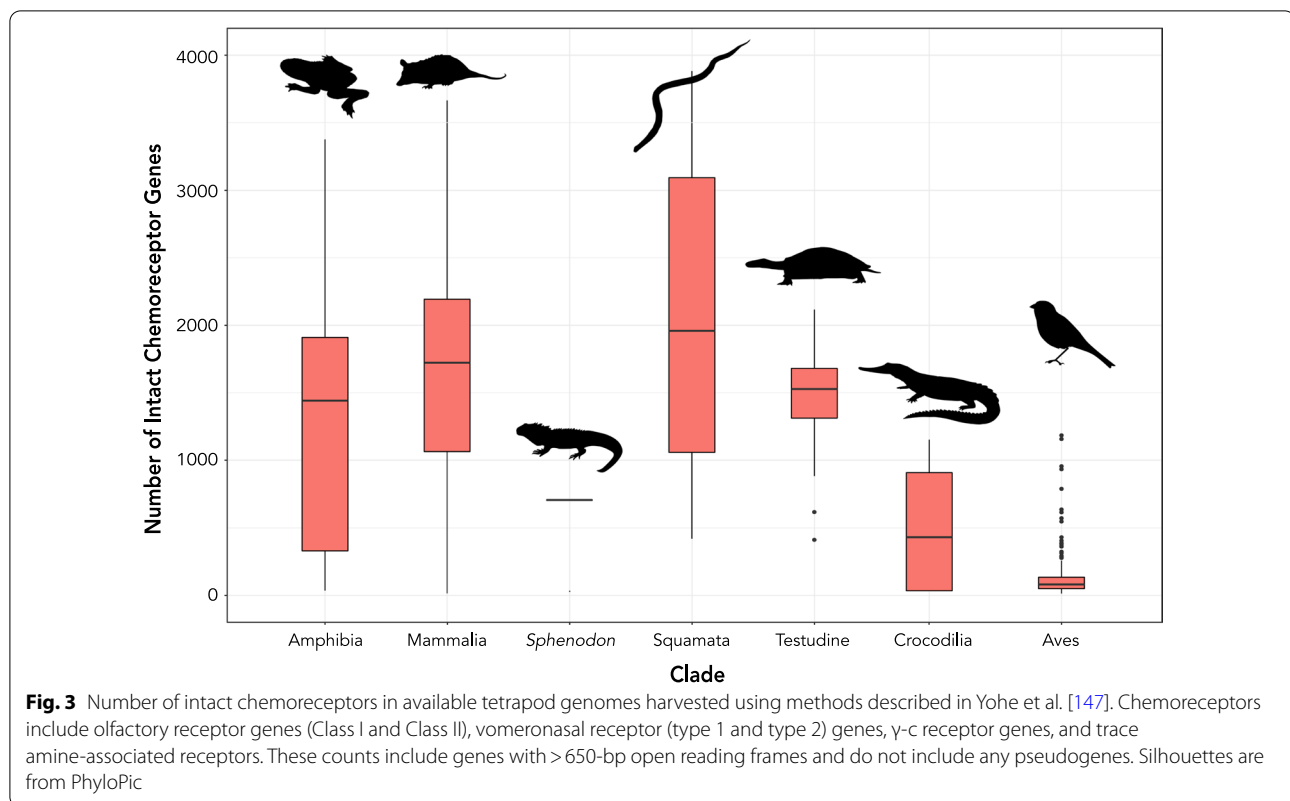
Chemosensory gene families are characterized by rapid evolutionary rates through high gene turnover and rapid diversification of homologous genes [146, 148–151]. The convergently evolved odorant receptor gene families in vertebrates and insects [146, 152, 153], for example, demonstrate some of the most extraordinary patterns of gene duplication and pseudogenization in animals, constantly expanding and contracting as species evolve through time [148]. We have made tremendous strides in our annotation and delimitation of human odorant receptors. However, placing these into the context of the vertebrate Tree of Life is challenged by their complex evolutionary history.

The expansion and contraction of the odorant receptor gene family have been linked to changes in the sensory ability of some animal lineages, including the adaptation to novel food resources [148, 154–157] or specializations in the social–chemical communication system [158, 159]. Accordingly, the hundreds/thousands of copies of distinct olfactory receptors likely reflect the diversity of odorant ligands that can be detected by the organism (Fig. 3, [144, 145]). Functional assays of olfactory receptor repertoires in a few model organisms suggest that odorant receptors act as labeled lines or in a combinatorial fashion, thereby enabling the specific detection and discrimination of chemicals and chemical mixtures vastly exceeding the number of receptor genes encoded in the genome, while

also being highly specific [160–167]. These observations have translational implications for understanding chemoperception at the population level, and future studies of the functional consequences of odorant receptor variation between individuals offer a particularly exciting avenue of future research.

The sheer enormity of these receptor gene repertoires (Fig. 3) calls into question the functional purpose of the extensive redundancy of these genes. Studies have repeatedly shown the elevated rates of evolution of chemoreceptor genes across animals, insects, and vertebrates alike [147, 155, 168]. Adaptive scenarios predict elevated rates of nonsynonymous substitutions over short lengths of evolutionary time (i.e., relative to rates of neutral synonymous substitutions) and would likely correspond to an environmental chemical signal or signals relevant to fitness. Evidence for adaptation is prevalent in *Drosophila*, where behavioral and biochemical assays have repeatedly demonstrated specialization of odorant genes in *Drosophila* species to ligands of host fruits (e.g., *Drosophila sechellia* and the host shift to morinda fruit) [169–173]. However, outside of model organisms, interpretations of correlative patterns have emerged but are still inconclusive. For example, orchid bee (Apidae: Euglossini) males collect specific organic chemical compounds from floral and non-floral sources available in their environment to create a perfume [174–176] released during a stereotypical display behavior at perching sites, and it is only in combination with perfume display that mating occurs [177–182]. Perfumes are most likely involved in sexual selection, presumably by enabling species-specific recognition or as an indicator of male fitness (Fig. 4A, [183, 184]). Homologous chemosensory receptor genes are highly differentiated and evolve under strong divergent selective pressures between even recently diverged (~150 kya) orchid bee lineages (Fig. 4B, C), suggesting adaptive differentiation of chemosensory genes [185, 186]. Although there is some evidence that divergence in perfume signals might be correlated with divergence in chemosensory genes [185], the link between genotypic and phenotypic evolution is still missing.

The scarcity of evidence of strong selection for direct receptor-odorant ligand binding suggests that the relationship between chemoreceptor genes and the environment is more complex than initially hypothesized. Rates of nonsynonymous substitutions for dozens of genes are consistent with diversifying selection across tetrapods [147, 168]. Substitution rates in the mouse lemur (*Microcebus*) pheromone receptors (vomeronasal type 1) are so high, and they occur at such exceptional copy numbers, that they have been described as genes on the verge of a “functional breakdown” [187]. Is it really possible that these dozens of chemoreceptor genes are solely under

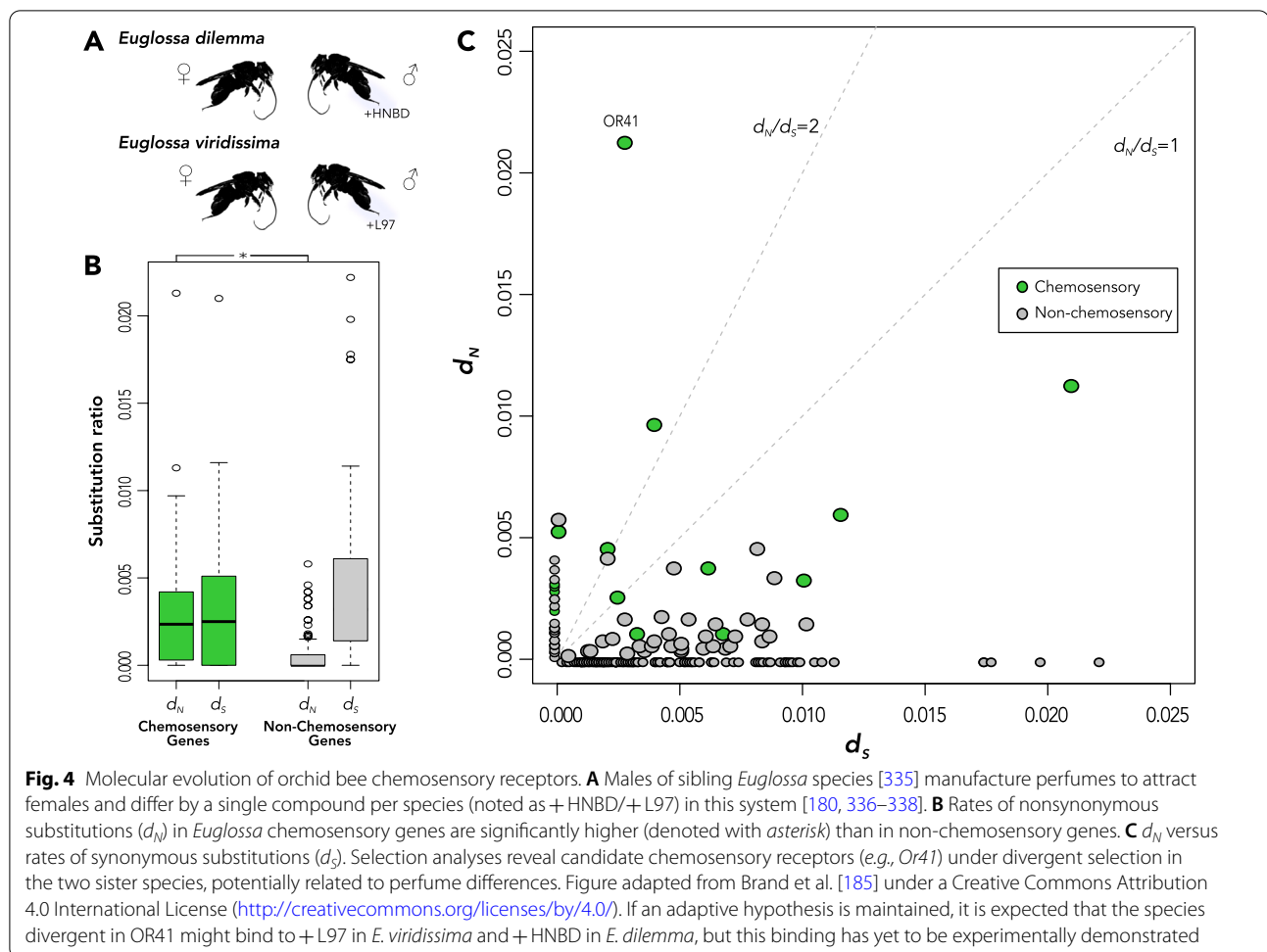


extensive positive selection? Neotropical bats illustrate an instance where differences in molecular rates are not directly explained by dietary ecology (e.g., plant-visiting versus animal feeding), but rather are associated with the chromosomal location of the genes [188]. Even within *Drosophila*, chemosensory genes have an elevated background of standing variation that may facilitate rapid adaptation in the event of shifts in ecological niches [189].

What is the evolutionary explanation of maintaining extensive copy numbers of highly variable but seemingly redundant genes in the genome? The answer may lie in navigating a complex world incorporating a highly dynamic chemical space [151, 190], as illustrated by mouse lemurs (*Microcebus*). In contrast to other primates (including humans) that typically possess few copies of pheromone receptor genes, these largely solitary primates possess orders of magnitude higher copy numbers of pheromone receptors (Fig. 5, [191]), and the primary form of conspecific communication in these arboreal and nocturnal primates occurs via pheromonal cues of scent-marking and anogenital dragging [192–194]. Females in estrous induce sexual behavior in males through proteins in their urine, and, in response, males establish their territories via urine scent-marking to signal dominance (Fig. 5A, [192, 195]). The high rates of molecular

evolution of chemoreceptors [187, 196] and rapid turnover of genes may facilitate sexual selection and species-specific adaptation [151] in this species. However, this adaptation does not come in the form of selection for a single variant binding to a specific ligand [187]. Rather, evolution of new receptors via gene duplication provides a genomic substrate to detect novel chemical cues signaled by competing males. Pseudogenization of receptors no longer relevant to the present environment leads to very low levels of orthology among closely related species (Fig. 5B, [197–199]). Mouse lemurs have speciated rapidly [200, 201]. Because receptors have continued to diversify among closely related species (Fig. 5C), it is a plausible hypothesis that increased redundancy and sequence variation among duplicates facilitate species diversification. This pattern is in sharp contrast to what is observed in bat vomeronasal receptors, in which there are very few copy number variants and a high retention of orthologs since bats diverged from ungulates and carnivores [202]. When these large gene families show low rates of gene retention at deep time scales, this may hint at conserved innate function.

Extrapolating function and adaptation across species, in particular for receptors, remains challenging. Identifying orthologous receptors among a pool of species- and individual-specific paralogs is not straightforward.



(See figure on next page.)

Fig. 5 Vomeroolfactory-mediated courtship and territoriality in mouse lemurs and the phylogenetic history of vomeronasal receptor type 1 (V1R) genes in primates. **A** The urine of the dominant male gray mouse lemur (*Microcebus murinus*) often contains a distinct steroid-like compound that suppresses reproductive behavior of other males, but it must stand out among competitors to attract females. *Dotted-lined arrows* indicate a weaker signal among the dominant male urine signal. **B** Gene tree of V1Rs in primates [187, 191], including the gray mouse lemur. *Black branches* indicate genes belonging to the mouse lemur, while *gray branches* belong to other primate groups. **C** V1R gene tree of lemurs [187], including several species of mouse and dwarf lemurs (Cheirogaleidae). *Black branches* are Cheirogaleidae and *gray branches* are other strepsirrhine primates. Sexual selection coupled with extensive gene duplication of vomeronasal receptors may have facilitated rapid speciation in Cheirogaleidae. Silhouettes were obtained from PhyloPic

Orthologs are usually assumed to have similar functions across species. However, orthologous olfactory receptors do not consistently share function [203]. In some cases, the same olfactory receptor in two species will respond to the same odorant ligand. In other cases, orthologs may differ in their responses to the same stimuli. Nevertheless, detecting adaptation in olfactory receptors is tractable with well-parameterized models

of molecular evolution and appropriate consideration of protein structure and function [204]. These receptors directly interact with environmental cues, with essential roles in the identification of stimuli that historically were directly related to fitness (e.g., finding food, mates, and predator avoidance). Comparative phylogenetic approaches will be critical for understanding this nexus of complex sensory-signaling and perception [190]. Given the immense representation of chemoreceptor genes in the human genome, understanding the

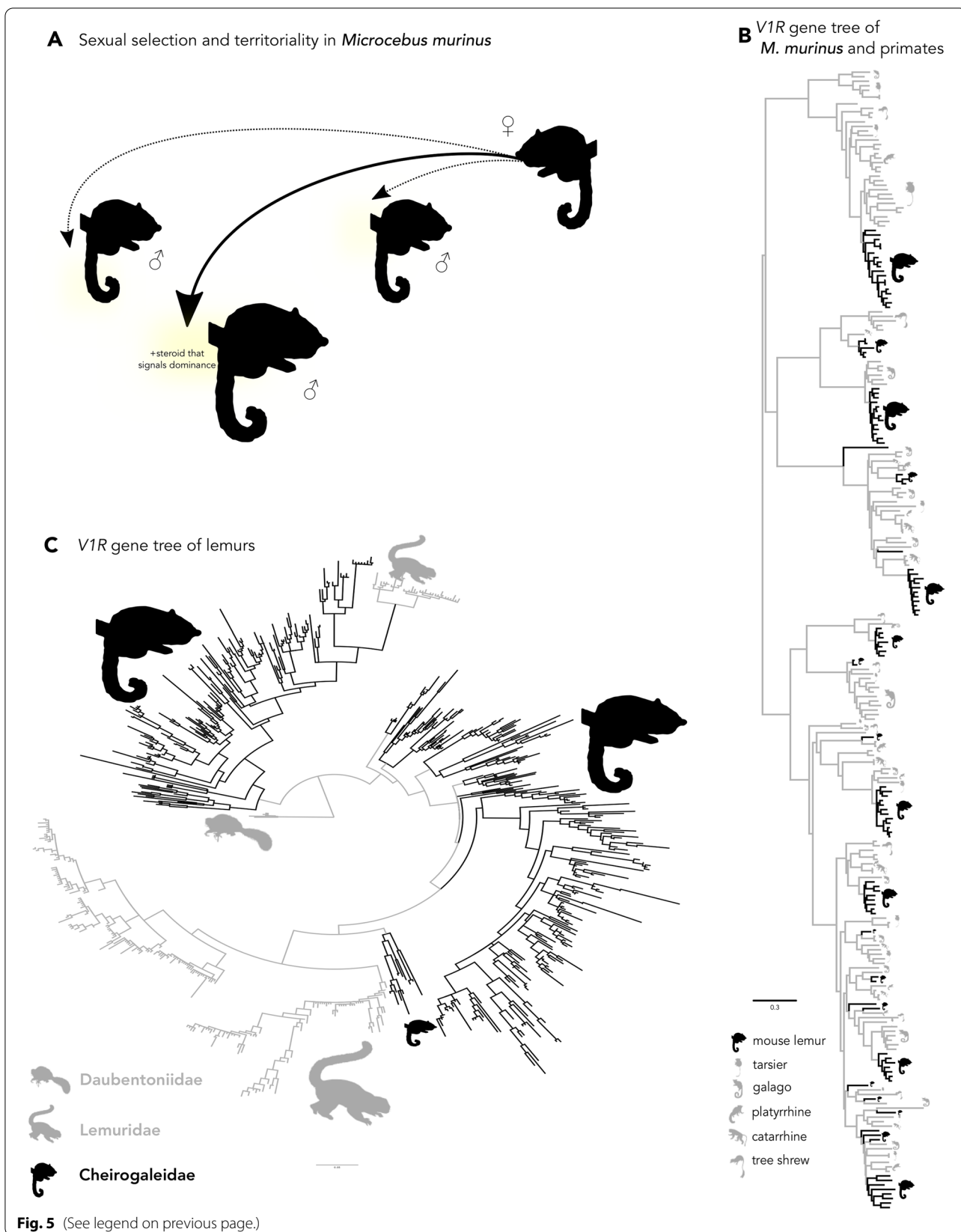


Fig. 5 (See legend on previous page.)

Table 1 Interspecific gene content variation of the KIR gene family

Species	# KIR genes in LRC	References
Homo sapiens	4–20	[29]
Chimpanzee	5–11	[281]
Orangutan	5–10	[282]
Gibbon	2–5	[283]
Baboon	>6	[284]
Macaque	4–17	[285]
Owl monkey	8	[286]
Gray mouse lemur	1	[287]
Mouse	0	[288]
Rat	1	[288]
Pig	1	[289]
Cow	18	[40]
Horse	1	[39]
Elephant	4	[38]
Dog	0	[49]
Cat	1 pseudogene	[49]
Sea lion and seal	1	[49]

Table 2 Interspecific gene content variation of the KLRA (Ly49) gene family

Species	# Ly49 genes in NKC	References
Homo sapiens	1 pseudogene	[290]
Chimpanzee	1 pseudogene	[284]
Orangutan	1	[291]
Gibbon	1	[284]
Baboon	1	[284]
African green monkey	1	[284]
Gray mouse lemur	1	[287]
Mouse	8–21	[35]
Rat	33–36	[35]
Pig	1	[292]
Cow	1	[293]
Horse	6	[294]
Dog	1	[49]
Cat	1	[292]
Sea lion and seal	1	[49]

complex evolutionary dynamics of these genes is crucial to understanding ourselves.

Breaking down the basis of enzymic metabolism

Enzymic metabolism genes have been speculated to evolve upon exposure to specific novel environmental pressures, including the ingestion of plants, fungi, lower eukaryotes, and bacteria. The importance of cytochrome P450 (CYP) genes in “animal–plant warfare”

was emphasized long ago [205], and the CYP2, CYP3, and CYP4 families encode enzymes that participate not only in metabolism of important endogenous substrates, but also in the metabolism of plant–fungal–bacterial–viral metabolites, drugs and other environmental pollutants, pigments, and biosynthesis of pheromones. Therefore, evolution and expansions of these gene families are extremely sensitive to changes in the environment of the organism, resulting in rapid up- and downregulation of gene expression. “Evolutionary blooms” would be expected in these three human P450 families and, indeed, they have been found [206]. Evidence of these (evolutionarily recent) blooms is presented by the high dissimilarity evident when comparing human and mouse CYP genes (Table 3).

The expansion of CYP genes reflects a larger phenomenon in which signals from the microbiome—which is commonly described as bacteria, fungi, viruses, and other microbes living synergistically in the digestive tract of the host [207, 208]—have had a profound effect on the evolution of host genomes. However, “digestive tract” signifies more than just “animals with stomachs”; guts have deep evolutionary origins. For example, the ambulacral groove of animals in the phylum Echinodermata or class Asterozoa and Edrioasteroidea extends from the mouth to the end of each ray or arm; each groove of each arm, in turn, has four rows of hollow tube feet that can be extended or withdrawn [209]. Even the cavitation of animals in the phylum Cnidaria (e.g., sponge, jellyfish, sea anemones and corals, etc.) functions in digestion of food and, thus, is likely to have a microbiome. Accordingly, it is likely that microbiome metabolites (as well as phytoplankton and other ancient simple plants) were among the first environmental signals to have been received by animal hosts and that their presence in turn generated selective pressure in the host to metabolize microbial metabolites.

Selective pressures on receptors of microbial metabolites—and the corresponding evolution of enzymic genes that interact with these receptors—likely shaped major trends in genome evolution [210, 211]. Hence, what we see today are multiple transcription factors (e.g., AHR, AHRR, ARNT, NR1I3 (CAR), PPARA, PPARG (PPAR-gamma), NFKB1/2, HNF4A/G, HNF1, NFE2L2 (Nrf2), NR1I2 (PXR), NR1H4 (FXR), ESR1/2, PGR, GHR, NPAS1, etc. [207, 212]), up- and down-regulating dozens or hundreds of genes that are members of enzymic metabolism gene families. As far as the CYP gene families (Table 3), transcription factors for the CYP1, 2, 3, and 4 families involved in plant metabolite and pheromone metabolism are distinctly different from the remaining fourteen P450 gene families (CYP5, 7, 8, 11, 17, 19, 20, 21, 24, 25, 26, 27, 39, and 46) that participate almost exclusively in endogenous

Table 3 Many of the heavily studied major classes of human enzymic metabolism genes

Mechanism	Encoded enzymes (gene root symbols, where applicable)	Families, classes, groups*	Representative reference(s)
<i>Oxidation</i>			
	Hydroxylases	Dozens of classes; > 1000 genes in genome	[295, 296]
	CYP monooxygenases (CYP; P450s)	Total of 18 families; 57 genes in human genome versus 102 in mouse <i>CYP1</i> (3 in <i>Homo</i> ; 3 in <i>Mus</i>) <i>CYP2</i> (16 in <i>Homo</i> ; 50 in <i>Mus</i>) <i>CYP3</i> (4 in <i>Homo</i> ; 9 in <i>Mus</i>) <i>CYP4</i> (12 in <i>Homo</i> ; 20 in <i>Mus</i>) <i>CYP5</i> , 7, 8, 11, 17, 19, 20, 21, 24, 26, 27, 39, 46, 51	[206, 213, 297] [206] [206] [206] [213]
	Flavin-dependent monooxygenases (FMO)	6 genes	[298, 299]
	Alcohol dehydrogenases (ADH)	8 genes	[300, 301]
	Aldehyde dehydrogenases (ALDH)	11 families, 4 subfamilies, 19 genes	[302, 303]
	Lysine (histone) demethylases (KDM)	8 groups; 24 genes	[304, 305]
	Monoamine oxidases (MAO)	Two genes	[306]
	Other amine oxidases	> 50 genes	[307]
	Peroxidases (PEX)	> 30 genes	[308, 309]
	Catalase	Three main groups; one gene in human	[310, 311]
	Oxidoreductases	< 40 genes, subunits	[312, 313]
<i>Reduction</i>			
	CYP reductive reactions (CYP)	CYP1, 2, 3, and 4 families	[314]
	NADPH-quinone oxidoreductases (NQO)	Two genes	[315]
	Aldo-keto reductases (AKR)	Two families, 7 subfamilies, 19 genes	[316]
	Carbonyl reductases (CBR)	Three genes	[316, 317]
	Short-chain dehydrogenases/reductases (SDR)	Up to 77 protein-coding genes	[318]
	Oxidoreductases	> 50	[318, 319]
<i>Hydrolysis</i>			
	Epoxide hydrolases (EPHX)	Up to 7 genes	[320, 321]
	Amidases	Three genes	[322]
	Esterases	At least 8 genes	[323, 324]
	Lipases (LIP)	22 genes	[325]
<i>Conjugation</i>			
	UDP glucuronosyltransferases (UGT)	4 families; 5 subfamilies; 23 genes†	[326, 327]
	Glutathione S-transferases (GST)	8 families; 21 genes	[328]
	Soluble sulfotransferases (SULT)	4 families; 9 subfamilies; 14 genes	[329]
	Membrane-bound sulfotransferases	37 genes	[330]
	<i>N</i> -acetyltransferases (NAT)	26 genes; two groups; GCN5-related (24 genes) and arylamine-related (2 genes)	[331, 332]
	Other transferases	Up to 41 groups; 265 protein-coding genes	HGNC

*Virtually all these enzyme classes, and members of each gene group with their chromosomal locations, can be found in genenames.org (HGNC)

†The “number of UGT genes” is debatable, because of the complex structures of the UGT1 and UGT2 loci

Details of the CYP gene family are perhaps the most studied on this list. The mammalian CYP2, CYP3, and CYP4 families each have many genes encoding plant-metabolizing enzymes, and the number of genes in each family varies widely among mammals [206]. The CYP1 family is involved in some plant-metabolizing enzymes but also arachidonic acid pathway-associated lipid mediator substrates [333], steroids, pigments, polycyclic aromatic hydrocarbons, polyhalogenated biphenyls, etc. The remaining 14 CYP families encode evolutionarily highly conserved enzymes with much more highly specific endogenous substrates (ablation of many or most of these is lethal), and the number of genes in each family is virtually identical among all mammalian genomes [213, 334]

critical life function pathways, instead of exogenous chemical metabolism [213]. This difference in function has opened new research opportunities. For example, mice having the ablation of all *Cyp2c* genes in a cluster

are completely viable, except for changes in metabolism for specific drugs or chemicals. As such, there is now a plan to create a methodically “humanized” mouse model for pharmacological and toxicological studies by

replacing all mouse *Cyp2*, *Cyp3*, and *Cyp4* genes with all human *CYP2*, *CYP3*, and *CYP4* genes [214]. Future work in other models and non-models that similarly target enzyme-metabolism-associated gene families could be of high translational relevance for human health.

Obtaining and maintaining universal gene family nomenclatures

The human gene mapping community began publishing proposals on standardized gene nomenclature in the 1970s “pre-genomics” era, e.g., [215, 216], and the HUGO Gene Nomenclature Committee (HGNC) has been naming human genes now for over 40 years [217]. However, these efforts had already been ongoing for mouse genetics since the 1940s [218]. Subsequent collaboration between the human and mouse gene nomenclature committee has enabled logical naming—and renaming—of many gene families to reflect evolutionary relationships. For example, the human CD300 gene family includes sequentially named genes that modulate a diverse array of immune responses, named CD300A–CD300H [219–221]. The original nomenclature for mouse models previously spanned a variety of root symbols including dendritic-cell-derived Ig-like receptor (DIgR), CRMF-35-like molecules (CLM), leukocyte mono-Ig-like receptors (LMIR), and myeloid-associated Ig-like receptors (MAIR) [219], but these mouse genes have since been renamed in line with the human genes using the Cd300 root. In more subtle cases, mismatches in nomenclature between closely related genes can be revealed through phylogenetic analyses; for instance, leukocyte receptor tyrosine kinase (LTK) in ray-finned fishes is a closer homolog of ALK receptor tyrosine kinase (ALK; previously anaplastic lymphoma kinase) in mammals than it is to LTK in mammals [74]. As sequencing efforts in models and non-models continue to accelerate, there is an urgent need to continue to standardize gene nomenclature not only within species, but also between species using phylogenetic criteria.

With regard to the standardization of nomenclature of gene families based on evolutionary divergence, the enzymic metabolism gene families offer one of the earliest examples, originating with the cytochrome P450 monooxygenase superfamily. From 1975 to 1985, leading scientists researching cytochrome P450 genes would convene at least once yearly, during the “Microsomes and Drug Oxidations” (MDO) or “P450 Biochemistry and Biophysics” symposia, to discuss what the best name might be for their (personal favorite) enzyme isolated in their respective laboratories. These enzymes (in human, rat, mouse, rabbit, pig, cow, chicken, fish, yeast, and *Pseudomonas putida*) are all membrane-bound (microsomal

or mitochondrial) proteins, except for the bacterial cytosolic enzyme which is soluble.

With the advent of isolating mRNA from ribosomes treated with antibodies to these detergent-solubilized enzymes, followed by reverse transcriptase to isolate the cDNA, the deduced amino acid sequence of these membrane-bound proteins could be determined. Once it was discovered that these P450 genes encoded a consensus sequence of eight amino acids in the enzyme active site—a consensus found in at least ten organisms as diverse as human and a bacterium [222]—the logical solution to standardized nomenclature was to base P450 gene names on evolutionary divergence.

At first, Roman numerals were included [222, 223], but this clumsy approach was quickly discarded. A root symbol “CYP” (standing for cytochrome P450) was then agreed upon, with Arabic numerals for gene families, capital letters for subfamilies, and Arabic numerals again for individual members; no subscripts or superscripts are allowed in standardized gene nomenclature, and hyphens are only used in specific exceptions and usually to separate two unrelated consecutive numerals. Two CYP genes encoding proteins (from the same species) that showed <40% amino acid sequence similarity were relegated to different gene families. Two CYP genes having ~40 to ~65% amino acid sequence similarity were assigned to the same family, whereas two genes encoding proteins that displayed $\geq 65\%$ amino acid sequence similarity were listed as members of the same subfamily [224]. Gene symbols are in all capital letters for human and most vertebrate genomes, i.e. *CYP1A2*, *CYP1B1*, *CYP17A1*, *CYP51A1*, etc. Mice deviate from this model due to historical contingency of an earlier nomenclature, capitalizing only the first letter; hence, the mouse orthologous genes are named *Cyp1a2*, *Cyp1b1*, *Cyp17a1*, *Cyp51a1*, etc.

In the process of naming the CYP gene families, they were originally somewhat arbitrarily divided into different classes of organisms (Table 4), based on an assumption that the number of P450 genes thought likely to exist in all animals on the planet would not exceed 50. However, this assumption has proven to be a striking underestimate. As of July 13, 2022, a total of 125,326 CYP genes in 8455 gene families among vertebrates, protozoa, plants, fungi, eubacteria, archaeobacteria, and viruses have been named [D R Nelson, personal communication]—with an anticipation that over one million will be reached. In general, plant genomes have far more P450 genes than animal genomes, because plant P450-mediated pathways are critical for virtually all life processes: growth, differentiation, defense (phytoalexin formation), fruit production, flower color, and formation of the attractive and repulsive scents of flowers [225].

Table 4 Historical format for the assignment of CYP gene families (<https://drnelson.uthsc.edu/>)

Gene family numbering	Class of organisms
CYP1–CYP49	Animals
CYP51–69	Yeast, Fungi
CYP71–99	Plants
CYP101–299	Bacteria, Viruses
CYP301–499	Animals
CYP501–699	Lower eukaryotes
CYP701–999	Plants
CYP1001–2999	Bacteria, Viruses
CYP3001–4999	Animals
CYP5001–6999	Lower eukaryotes
CYP7001–9999	Plants
CYP10001–29,999	Bacteria, Viruses*
CYP30001–49,999	(Animals)*
CYP50001–69,999	Lower eukaryotes

*No genes yet included in these ranges

Gene family numbers for other P450 genes were assigned in chronological sequence as they were identified (hence, CYP6, 9, 12 are insect gene families; *CYP10A1* is a pond snail gene; and *cyp-13A1* is a nematode gene, etc.). As a consequence, human CYP gene families are not named with sequential numbers, but rather CYP1, 2, 3, 4, 5, 7, 8, 11, 17, 19, 20, 21, 24, 25, 26, 27, 39, and 46 [213]. Although these authoritative efforts have yielded a stable nomenclature for this gene family, they also reflect a problem of sustainability. The herculean task of continually updating and maintaining a stable nomenclature of CYP genes—and specifically P450 genes—has been undertaken by a single curator, David R Nelson. As new genomes are published, the extraction and naming of all present P450 genes using BLAST searches (by DR Nelson) remains the predominant source of nomenclature updates for non-models. Dependency of nomenclature on single individuals augurs gene identification crises like the taxonomic identification crises lamented by current-day systematics [226]. Sustainable solutions to gene nomenclature are achievable through sustained funding and collaboration between gene nomenclature committees across multiple organisms and represent an exciting challenge and research opportunity that can integrate the expertise of gene taxonomists and computational biologists, aspirationally generating a well-defined and extensible nomenclature for comparative genomics.

As the CYP gene superfamily was developed during the late 1980s and early 1990s, dozens of other enzymatic metabolism gene families (e.g., Table 3) also began to establish standardized nomenclature systems based

on evolutionary divergence. Work on these and many other gene families such as the olfactory receptors [227] by the HGNC and more recently also by the Vertebrate Gene Nomenclature Committee (VGNC, vertebrate.genenames.org), a sister project of the HGNC [228], has enabled the implementation of a systematic approach to standardized nomenclature based on naming paralogous genes originating from a common ancestor. However, the proliferation of genomes from non-model species now provides an opportunity to stabilize a consistent gene nomenclature at deeper taxonomic scales, as has already been instigated for cytochrome P450 genes, through collaboration between established nomenclature committees [229]. Other efforts have ensured that the immunoglobulin (*Ig*) genes that encode antibodies across ray-finned fishes were recently standardized, because *IgT* and *IgZ* were found to be evolutionary forms of the same antibody that was independently discovered and named in different species [230]. This standardization across half of all living vertebrates demonstrates the utility of phylogenetic comparative genomics to gene nomenclature at such large taxonomic scales. As genome sequences continue to accrue, gene taxonomists are now poised to work with nomenclature committees to harness the power of this comparative framework and decode the history of paralog diversification in complex gene families, thereby stabilizing nomenclature across the Tree of Life.

Veils of deep ancestry that limit our perception of gene family evolution

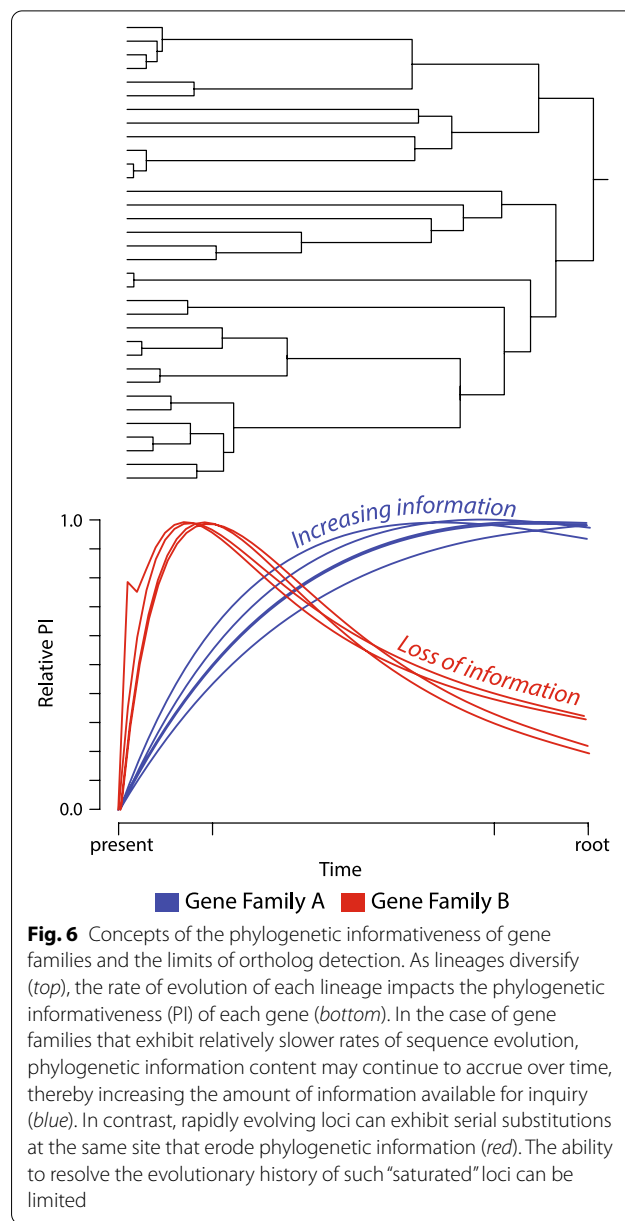
In general, investigations of gene families are rooted in comparative approaches. As genes of interest become identified, approaches based on sequence similarity or homology are used to detect similar genes within a target genome (intragenomically or intergenomically). Some of the most commonly used sequence similarity tools include BLAST [231, 232], PSI-BLAST [233, 234], diamond [235], and HMMER3 [236, 237] that collectively represent routine aspects of bioinformatic pipelines. Heuristic searches such as BLAST perform global and local alignments using dynamic programming algorithms (e.g., Needleman–Wunsch and Smith–Waterman algorithms) that require a database such as those hosted by NCBI, EMBL, or Ensembl for searching. The results are obtained based on a predetermined *e* value threshold, and the most similar matches are selected. However, *e* value-based metrics are not infallible, because they can give different results based on the completeness of the reference database. Similarity methods that apply *e* values are also vulnerable to inflating relatedness, as it is possible that only a small fraction of the query sequences is being considered in the analysis, rather than the whole sequence. This problem arises frequently in

the annotation of non-model species when using highly curated databases as a reference (e.g., UniProtKB [238]). To complement *e* value-based identification, the identity of gene family members is often validated through phylogenetic analyses such as those conducted in IQTREE [239] or RevBayes [240].

Phylogenetic analyses allow candidate genes to be placed among known reference sequences to verify identification, assess the evolutionary history of gene gains and losses, and delimit unique gene expansions within lineages [59, 74]. However, it is well known that some genes are more amenable to phylogenetic inference than others [241–245], and the utility of each locus for evolutionary inference is dependent on the phylogenetic hypotheses being addressed [246–248]. Much of the heterogeneous utility of loci for phylogenetics can be attributed to variance in the rate of molecular evolution between sites, wherein fast-evolving sites can erode phylogenetic information [249, 250]. Just as phylogenetic information can be eroded by evolution, so can identification of homology [251]. In the presence of high rates of sequence evolution, “phylogenetic noise” in the data can mask or even mislead interpretation of the relationship of the queried gene family with the newly detected candidate genes in the database [251]. This problem arises as a natural consequence of numerous substitutions occurring in a gene over time and has the potential to promote erroneous inference [249, 252–254]. In cases with little true signal of evolutionary history, this phylogenetic noise can accumulate, leading to the estimation of phylogenetic tree topologies that appear to have strong statistical support [255–257]. These problems are not new to phylogenetics, and solutions and approaches to their mitigation should be adopted for studies of gene family evolution.

The point at which noise (molecular changes reflecting chance convergence or parallelism) erodes signal (differences in nucleotides or amino acids since common ancestry) is not commonly accounted for—as an issue of statistical power—when defining the origins of genes. High rates of evolution at sites—especially genome-wide—will increase genetic distances between loci at deeper timescales [250, 258]. Consequently, an inability to determine which sites are evolving at rates predicted to contribute signal versus noise poses a major impediment in the selection of orthologs (Fig. 6). Common practices, such as using *e* value thresholds, do not address this problem [251]. Rates of evolution can vary between genes and lineages, and therefore cutoffs for accurate identification of orthologs differ between organismal groups and time scales [259].

Once genes become highly divergent, it becomes challenging—and then impossible—to assign homology of



molecular sequence characters. However, quantifications of phylogenetic information loss are rarely incorporated into studies of homology. Quantification of evolutionary information loss through time requires consideration of the information gained by adding genetically distant taxa [250, 260, 261], characteristics of the sequence data (e.g., biases in nucleotide compositions [262, 263], amino acid versus nucleotide data [264], etc.), as well as molecular rate heterogeneity between taxa or loci [250, 255, 265, 266]. Such an integration facilitates calculation of the accumulation of signal versus loss of information due to chance convergences or parallelisms that provide

an expectation of where power is highest for phylogenetic inference [246, 250, 267]. What is now needed is a mathematical framework that builds on such theory from phylogenetics, to predict when the inability to detect an ortholog would stem from a lack of phylogenetic information versus a likely true absence of an ortholog. This kind of framework will illuminate the predicted limits of ortholog detection that are critical to establishing where we are, and are not, confident in identifying the origins of genes in the human genome.

The future of human genomics is comparative

Central to the diversification of the human genome are duplications of genes or entire genomic regions. Over short evolutionary timescales, gene birth events (i.e., gene duplications that create paralogs) may provide redundancy for essential genes or increased levels of a gene product [268]. Over longer stretches of evolutionary time, these gene birth events provide the molecular basis for sequence evolution through neofunctionalization (i.e., the acquisition of new functions), subfunctionalization (all functions of the original gene are maintained, but divided between the gene copies), or non-functionalization (accumulation of deleterious mutations resulting in gene death) [269, 270]. The function of specific gene duplications and clusters has been well studied in select species [55, 271]. However, the potential role of rapid paralog evolution in gene families as the substrate for further genomic novelty is only beginning to be explored. Testing for shifts in selection that have promoted the generation of new phenotypes in response to ecological opportunity is a fundamental aspect of macroevolution [272–274]. There is little doubt that the rapid expansion of a gene family can provide the necessary genomic foundation for pulses of phenotypic evolution. What remains unclear are the relationships between ecological opportunity, changes in selection, gene birth–death dynamics, and the timing of phenotypic diversification in the history of a lineage. Recent work has highlighted that large-scale genomic changes are temporally decoupled from the onset of phenotypic diversification within both the iconic adaptive radiations of African rift lake cichlids and Antarctic notothenioids [275, 276]. These findings highlight the critical, but often neglected, role of historical contingency in understanding adaptive evolution [277]. Therefore, as the taxonomic diversity of available genomes continues to increase, comparative studies with a wider taxonomic scope will be critical to determine the relationship between the tempo of gene family evolution and present-day phenotypes.

Successive tandem gene duplication events are of particular interest in clustered gene families. Relative

to surrounding regions, clustered gene families often display disproportionately high levels of sequence and gene copy diversity. As such, clustered gene families may be hotspots for additional gene birth events, higher rates of gene death (loss), nucleotide substitution, sharing of regulatory sequences among multiple genes in tandem, exon swapping, and interlocus gene conversion. However, the diversification dynamics of gene clusters across vertebrate macroevolutionary history remains largely unexplored. What is needed is the continued development of theory [278] and novel tools [279, 280] that can be used to generate a unified paradigm for understanding the dynamics of clustered gene diversification both within and between species. As diversity within gene clusters contributes to a range of genetic disorders, such a paradigm is central to linking hyper-diverse clusters between model organisms and humans and might drive both diagnostic advances and the development of novel therapeutics.

Acknowledgements

We thank Philipp Brand as well as members of the Dornburg, Yoder, and Townsend lab groups for helpful discussions that shaped the content of this review.

Author contributions

AD, JPT, EB, and VV conceived of this review. AD and JPT designed the manuscript, coordinated writing efforts, and worked with authors on all sections of this manuscript. For initial drafts of each section, AD, EB, and ZW wrote the vision portion, MB wrote the gene expression and environmental change portions, AD and JAY wrote the immunogenetics portion, AD, RM, and JAY wrote the cancer genes portion, LRY wrote the chemosensory portion, DWN and EB wrote the enzymic metabolism and gene nomenclature portions, and AD and JPT wrote the introduction and final two sections of the manuscript. JAY completed Figs. 1 and 2, LRY generated Figs. 3, 4, 5, and AD made Fig. 6. JAY created Tables 1 and 2 and DWN created Tables 3 and 4. All authors edited the manuscript. All authors read and approved the final manuscript.

Funding

This review was supported, in part, by grants from the National Science Foundation (IOS-1755242 to AD; IOS-1755330 to JAY; DEB-2032073 to LRY; IOS-1916137 to JPT); NIH P30 ES006096 to DWN; NIH P42ES033815-01 to VV, and National Human Genome Research Institute (NHGRI) grant U24HG003345 and Wellcome grant 208349/Z/17/Z to EB. The funding bodies played no role in writing the manuscript.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Bioinformatics and Genomics, UNC-Charlotte, Charlotte, NC, USA. ²Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA. ³Department of Biological Sciences, College of Science and Mathematics, Auburn University, Auburn, AL, USA. ⁴Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA. ⁵Department of Haematology, University of Cambridge School of Clinical Medicine, Cambridge, UK. ⁶European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK. ⁷Department of Environmental Health, Center for Environmental Genetics, University of Cincinnati Medical Center, P.O. Box 670056, Cincinnati, OH 45267, USA. ⁸Department of Pediatrics and Molecular Developmental Biology, Division of Human Genetics, Cincinnati Children's Hospital, Cincinnati, OH 45229, USA. ⁹Department of Molecular Biomedical Sciences,

College of Veterinary Medicine, North Carolina State University, Raleigh, NC, USA. ¹⁰Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA.

Received: 28 July 2022 Accepted: 12 October 2022
Published online: 11 November 2022

References

- Yoder JA, Litman GW. The phylogenetic origins of natural killer receptors and recognition: relationships, possibilities, and realities. *Immunogenetics*. 2011;63:123–41.
- Flajnik MF, Du Pasquier L. Evolution of innate and adaptive immunity: can we draw a line? *Trends Immunol*. 2004;25:640–4.
- Tassia MG, Whelan NV, Halanych KM. Toll-like receptor pathway evolution in deuterostomes. *Proc Natl Acad Sci USA*. 2017;114:7055–60.
- Grimholt U, Tsukamoto K, Azuma T, Leong J, Koop BF, Dijkstra JM. A comprehensive analysis of teleost MHC class I sequences. *BMC Evol Biol*. 2015;15:32.
- Schartl M. Beyond the zebrafish: diverse fish species for modeling human disease. *Dis Model Mech*. 2014;7:181–92.
- Yohe LR, Liu L, Dávalos LM, Liberles DA. Protocols for the molecular evolutionary analysis of membrane protein gene duplicates. *Methods Mol Biol*. 2019;1851:49–62.
- Gu X, Zhang Z, Huang W. Rapid evolution of expression and regulatory divergences after yeast gene duplication. *Proc Natl Acad Sci USA*. 2005;102:707–12.
- Trail F, Wang Z, Stefanko K, Cubba C, Townsend JP. The ancestral levels of transcription and the evolution of sexual phenotypes in filamentous fungi. *PLoS Genet*. 2017;13:e1006867.
- Whitehead A, Crawford DL. Variation within and among species in gene expression: raw material for evolution. *Mol Ecol*. 2006;15:1197–211.
- Rohlf RV, Harrigan P, Nielsen R. Modeling gene expression evolution with an extended ornstein-uhlenbeck process accounting for within-species variation. *Mol Biol Evol*. 2014. <https://doi.org/10.1093/molbev/mst190>.
- Loehlin DW, Carroll SB. Expression of tandem gene duplicates is often greater than twofold. *Proc Natl Acad Sci USA*. 2016;113:5988–92.
- Rohlf RV, Nielsen R. Phylogenetic ANOVA: the expression variance and evolution model for quantitative trait evolution. *Syst Biol*. 2015;64:695–708.
- Wang Z, Gudibanda A, Ugwuowo U, Trail F, Townsend JP. Using evolutionary genomics, transcriptomics, and systems biology to reveal gene networks underlying fungal development. *Fungal Biol Rev*. 2018;32:249–64.
- Hodgins-Davis A, Rice DP, Townsend JP. Gene expression evolves under a house-of-cards model of stabilizing selection. *Mol Biol Evol*. 2015. <https://doi.org/10.1093/molbev/msv094>.
- Lemos B, Meiklejohn CD, Cáceres M, Hartl DL. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution*. 2005;59:126–37.
- Metzger BPH, Duveau F, Yuan DC, Tryban S, Yang B, Wittkopp PJ. Contrasting frequencies and effects of cis- and trans-regulatory mutations affecting gene expression. *Mol Biol Evol*. 2016;33:1131–46.
- Bedford T, Hartl DL. Optimization of gene expression by natural selection. *Proc Natl Acad Sci USA*. 2009;106:1133–8.
- Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet*. 2011;13:59–69.
- Labbé P, Milesi P, Yébakima A, Pasteur N, Weill M, Lenormand T. GENE-dosage effects on fitness in recent adaptive duplications: ace-1 in the mosquito *Culex pipiens*. *Evolution*. 2014;68:2092–101.
- Nelson JA. Oxygen consumption rate v. rate of energy utilization of fishes: a comparison and brief history of the two measurements. *J Fish Biol*. 2016;88:10–25.
- Pörtner H-O, Bock C, Mark FC. Oxygen- and capacity-limited thermal tolerance: bridging ecology and physiology. *J Exp Biol*. 2017;220:2685–96.
- Bernal MA, Donelson JM, Veilleux HD, Ryu T, Munday PL, Ravasi T. Phenotypic and molecular consequences of stepwise temperature increase across generations in a coral reef fish. *Mol Ecol*. 2018. <https://doi.org/10.1111/mec.14884>.
- Bernal MA, Schunter C, Lehmann R, Lightfoot DJ, Allan BJM, Veilleux HD, et al. Species-specific molecular responses of wild coral reef fishes during a marine heatwave. *Sci Adv*. 2020;6:eaay3423.
- Bernal MA, Ravasi T, Rodgers GG, Munday PL, Donelson JM. Plasticity to ocean warming is influenced by transgenerational, reproductive, and developmental exposure in a coral reef fish. *Evol Appl*. 2022. <https://doi.org/10.1111/eva.13337>.
- Alrafiah A, Karyka E, Coldicott I, Iremonger K, Lewis KE, Ning K, et al. Plastin 3 promotes motor neuron axonal growth and extends survival in a mouse model of spinal muscular atrophy. *Mol Ther Methods Clin Dev*. 2018. <https://doi.org/10.1016/j.omtm.2018.01.007>.
- Bernal MA, Schmidt E, Donelson JM, Munday PL, Ravasi T. Molecular response of the brain to cross-generational warming in a coral reef fish. *Front Mar Sci*. 2022. <https://doi.org/10.3389/fmars.2022.784418>.
- Buckingham LJ, Ashby B. Coevolutionary theory of hosts and parasites. *J Evol Biol*. 2022;35:205–24.
- Middleton D, Gonzelez F. The extensive polymorphism of KIR genes. *Immunology*. 2010;129:8–19.
- Trowsdale J, Jones DC, Barrow AD, Traherne JA. Surveillance of cell and tissue perturbation by receptors in the LRC. *Immunol Rev*. 2015;267:117–36.
- Pelak K, Need AC, Fellay J, Shianna KV, Feng S, Urban TJ, et al. Copy number variation of KIR genes influences HIV-1 control. *PLoS Biol*. 2011;9:e1001208.
- Tukwasibwe S, Nakimuli A, Traherne J, Chazara O, Jayaraman J, Trowsdale J, et al. Variations in killer-cell immunoglobulin-like receptor and human leukocyte antigen genes and immunity to malaria. *Cell Mol Immunol*. 2020. <https://doi.org/10.1038/s41423-020-0482-z>.
- Sorgho PA, Djigma FW, Martinson JJ, Yonli AT, Nagalo BM, Compaore TR, et al. Role of Killer cell immunoglobulin-like receptors (KIR) genes in stages of HIV-1 infection among patients from Burkina Faso. *Biomol Concepts*. 2019;10:226–36.
- Agrawal S, Prakash S. Significance of KIR like natural killer cell receptors in autoimmune disorders. *Clin Immunol*. 2020;216:108449.
- Mansouri M, Villard J, Ramzi M, Alavianmehr A, Farjadian S. Impact of donor KIRs and recipient KIR/HLA class I combinations on GVHD in patients with acute leukemia after HLA-matched sibling HSCT. *Hum Immunol*. 2020;81:285–92.
- Rahim MMA, Makrigiannis AP. Ly49 receptors: evolution, genetic diversity, and impact on immunity. *Immunol Rev*. 2015;267:137–47.
- Lee SH, Girard S, Macina D, Busà M, Zafer A, Belouchi A, et al. Susceptibility to mouse cytomegalovirus is associated with deletion of an activating natural killer cell receptor of the C-type lectin superfamily. *Nat Genet*. 2001;28:42–5.
- Parham P, Norman PJ, Abi-Rached L, Guethlein LA. Human-specific evolution of killer cell immunoglobulin-like receptor recognition of major histocompatibility complex class I molecules. *Philos Trans R Soc Lond B Biol Sci*. 2012;367:800–11.
- Gusel'nikov SV, Taranin AV. Unraveling the LRC evolution in mammals: IGSF1 and A1BG provide the keys. *Genome Biol Evol*. 2019;11:1586–601.
- Futas J, Horin P. Natural killer cell receptor genes in the family equidae: not only Ly49. *PLoS ONE*. 2013;8:e64736.
- Sanderson ND, Norman PJ, Guethlein LA, Ellis SA, Williams C, Breen M, et al. Definition of the cattle killer cell Ig-like receptor gene family: comparison with aurochs and human counterparts. *J Immunol*. 2014;193:6016–30.
- Barrow AD, Trowsdale J. The extended human leukocyte receptor complex: diverse ways of modulating immune responses. *Immunol Rev*. 2008;224:98–123.
- Martin AM, Kulski JK, Witt C, Pontarotti P, Christiansen FT. Leukocyte Ig-like receptor complex (LRC) in mice and men. *Trends Immunol*. 2002;23:81–8.
- Hudson LE, Allen RL. Leukocyte Ig-like receptors: a model for MHC class I disease associations. *Front Immunol*. 2016;7:281.
- Hogan L, Bhuju S, Jones DC, Laing K, Trowsdale J, Butcher P, et al. Characterisation of bovine leukocyte Ig-like receptors. *PLoS ONE*. 2012;7:e34291.

45. Takai T. Paired immunoglobulin-like receptors and their MHC class I recognition. *Immunology*. 2005;115:433–40.
46. Tun T, Kubagawa Y, Dennis G, Burrows PD, Cooper MD, Kubagawa H. Genomic structure of mouse PIR-A6, an activating member of the paired immunoglobulin-like receptor gene family. *Tissue Antigens*. 2003;61:220–30.
47. Schwartz JC, Hammond JA. The unique evolution of the pig LRC, a single KIR but expansion of LILR and a novel Ig receptor family. *Immunogenetics*. 2018;70:661–9.
48. Schenkel AR, Kingry LC, Slayden RA. The Ly49 gene family: a brief guide to the nomenclature, genetics, and role in intracellular infection. *Front Immunol*. 2013;4:90.
49. Hammond JA, Guethlein LA, Abi-Rached L, Moesta AK, Parham P. Evolution and survival of marine carnivores did not require a diversity of killer cell Ig-like receptors or Ly49 NK cell receptors. *J Immunol*. 2009;182:3618–27.
50. Rojo S, Burshtyn DN, Long EO, Wagtmann N. Type I transmembrane receptor with inhibitory function in mouse mast cells and NK cells. *J Immunol*. 1997;158:9–12.
51. Wang LL, Mehta IK, LeBlanc PA, Yokoyama WM. Mouse natural killer cells express gp49B1, a structural homologue of human killer inhibitory receptors. *J Immunol*. 1997;158:13–7.
52. Shen L, Stuge TB, Zhou H, Khayat M, Barker KS, Quiniou SMA, et al. Channel catfish cytotoxic cells: a mini-review. *Dev Comp Immunol*. 2002;26:141–9.
53. Fischer U, Koppang EO, Nakanishi T, Teleost T and NK cell immunity. *Fish Shellfish Immunol*. 2013;35:197–206.
54. Litman GW, Hawke NA, Yoder JA. Novel immune-type receptor genes. *Immunol Rev*. 2001;181:250–9.
55. Yoder JA. Form, function and phylogenetics of NITRs in bony fish. *Dev Comp Immunol*. 2009;33:135–44.
56. Cannon JP, Haire RN, Magis AT, Eason DD, Winfrey KN, Hernandez Prada JA, et al. A bony fish immunological receptor of the NITR multigene family mediates allogeneic recognition. *Immunity*. 2008;29:228–37.
57. Wei S, Zhou J-M, Chen X, Shah RN, Liu J, Orcutt TM, et al. The zebrafish activating immune receptor Nit9 signals via Dap12. *Immunogenetics*. 2007;59:813–21.
58. Traver D, Yoder JA. Chapter 19: immunology. In: Cartner SC, Eisen JS, Farmer SC, Guillemain KJ, Kent ML, Sanders GE, editors. *The zebrafish in biomedical research*. Academic Press; 2020. p. 191–216.
59. Dornburg A, Wcisel DJ, Zapfe K, Ferraro E, Roupe-Abrams L, Thompson AW, et al. Holosteans contextualize the role of the teleost genome duplication in promoting the rise of evolutionary novelties in the ray-finned fish innate immune system. <https://doi.org/10.1101/2021.06.11.448072>
60. Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet*. 2016;48:427–37.
61. Thompson A, Hawkins M, Parey E, Wcisel D, Ota T, Kawasaki K, et al. The genome of the bowfin (*Amia calva*) illuminates the developmental evolution of ray-finned fishes. <https://doi.org/10.21203/rs.3.rs-92055/v1>
62. Wcisel DJ, Ota T, Litman GW, Yoder JA. Spotted gar and the evolution of innate immune receptors. *J Exp Zool B Mol Dev Evol*. 2017;328:666–84.
63. Wcisel DJ, Yoder JA. The confounding complexity of innate immune receptors within and between teleost species. *Fish Shellfish Immunol*. 2016;53:24–34.
64. Rodríguez-Núñez I, Wcisel DJ, Litman GW, Yoder JA. Multigene families of immunoglobulin domain-containing innate immune receptors in zebrafish: deciphering the differences. *Dev Comp Immunol*. 2014;46:24–34.
65. Wcisel DJ, Dornburg A, McConnell SC, Hernandez KM, Andrade J, de Jong JLO, et al. A highly diverse set of novel immunoglobulin-like transcript (NILT) genes in zebrafish indicates a wide range of functions with complex relationships to mammalian receptors. *Cold Spring Harbor Laboratory*; 2022; <https://doi.org/10.1101/2022.04.21.489081.abstract>
66. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
67. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011. <https://doi.org/10.1016/j.cell.2011.02.013>.
68. Haupt S, Haupt Y. P53 at the start of the 21st century: lessons from elephants. *F1000Res*. 2017;6:2041.
69. Belyi VA, Ak P, Markert E, Wang H, Hu W, Puzio-Kuter A, et al. The origins and evolution of the p53 family of genes. *Cold Spring Harb Perspect Biol*. 2010;2:a001198.
70. Nunney L. Size matters: height, cell number and a person's risk of cancer. *Proc Biol Sci*. 2018. <https://doi.org/10.1098/rspb.2018.1743>.
71. Casás-Selves M, Degregori J. How cancer shapes evolution, and how evolution shapes cancer. *Evolution*. 2011;4:624–34.
72. White MC, Holman DM, Boehm JE, Peipins LA, Grossman M, Henley SJ. Age and cancer risk: a potentially modifiable relationship. *Am J Prev Med*. 2014;46:57–15.
73. Vazquez JM, Sulak M, Chigurupati S, Lynch VJ. A zombie LIF Gene in elephants is upregulated by TP53 to Induce apoptosis in response to DNA damage. *Cell Rep*. 2018;24:1765–76.
74. Dornburg A, Wang Z, Wang J, Mo ES, López-Giráldez F, Townsend JP. Comparative genomics within and across bilaterians illuminates the evolutionary history of ALK and LTK proto-oncogene origination and diversification. *Genome Biol Evol*. 2021. <https://doi.org/10.1093/gbe/evaa228>.
75. De Munck S, Provost M, Kurikawa M, Omori I, Mukohyama J, Felix J, et al. Structural basis of cytokine-mediated activation of ALK family receptors. *Nature*. 2021;600:143–7.
76. Hallberg B, Palmer RH. Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. *Nat Rev Cancer*. 2013;13:685–700.
77. Janostiak R, Malvi P, Wajapeyee N. Anaplastic lymphoma kinase confers resistance to BRAF kinase inhibitors in melanoma. *iScience*. 2019;16:453–67.
78. Katayama R. Resistance to anaplastic lymphoma kinase (ALK) tyrosine kinase inhibitors (TKIs) in patients with lung cancer: single mutations, compound mutations, and other mechanisms of drug resistance. *Ther Strateg Overcome ALK Resist Cancer*. 2021. <https://doi.org/10.1016/b978-0-12-821774-0.00015-2>.
79. Englund C, Lorén CE, Grabbe C, Varshney GK, Deleuil F, Hallberg B, et al. Jeb signals through the ALK receptor tyrosine kinase to drive visceral muscle fusion. *Nature*. 2003;425:512–6.
80. Ishihara T, Iino Y, Mohri A, Mori I, Gengyo-Ando K, Mitani S, et al. HEN-1, a secretory protein with an LDL receptor motif, regulates sensory integration and learning in *Caenorhabditis elegans*. *Cell*. 2002;109:639–49.
81. Reshetnyak AV, Murray PB, Shi X, Mo ES, Mohanty J, Tome F, et al. Augmentor α and β (FAM150) are ligands of the receptor tyrosine kinases ALK and LTK: hierarchy and specificity of ligand–receptor interactions. *Proc Natl Acad Sci*. 2015. <https://doi.org/10.1073/pnas.1520099112>.
82. Mo ES, Cheng Q, Reshetnyak AV, Schlessinger J, Nicoli S. Alk and Ltk ligands are essential for iridophore development in zebrafish mediated by the receptor tyrosine kinase Ltk. *Proc Natl Acad Sci*. 2017. <https://doi.org/10.1073/pnas.1710254114>.
83. Barclay AN, Brown MH. The SIRP family of receptors and immune regulation. *Nat Rev Immunol*. 2006;6:457–64.
84. Murata Y, Saito Y, Kotani T, Matozaki T. CD47-signal regulatory protein α signaling system and its application to cancer immunotherapy. *Cancer Sci*. 2018;109:2349–57.
85. Dornburg A, Yoder JA. On the relationship between extant innate immune receptors and the evolutionary origins of jawed vertebrate adaptive immunity. *Immunogenetics*. 2022. <https://doi.org/10.1007/s00251-021-01232-7>.
86. van Beek EM, Cochrane F, Barclay AN, van den Berg TK. Signal regulatory proteins in the immune system. *J Immunol*. 2005;175:7781–7.
87. Ichigotani Y, Matsuda S, Machida K, Oshima K, Iwamoto T, Yamaki K, et al. Molecular cloning of a novel human gene (SIRP-B2) which encodes a new member of the SIRP/SHPS-1 protein family. *J Hum Genet*. 2000;45:378–82.
88. Viertlboeck BC, Schmitt R, Göbel TW. The chicken immunoregulatory receptor families SIRP, TREM, and CMRF35/CD300L. *Immunogenetics*. 2006;58:180–90.
89. Matlung HL, Szilagy K, Barclay NA, van den Berg TK. The CD47-SIRP α signaling axis as an innate immune checkpoint in cancer. *Immunol Rev*. 2017;276:145–64.
90. Liu L, Xiang Y-R. “Eating” Cancer cells by blocking CD47 signaling: Cancer therapy by targeting the innate immune checkpoint. *Cancer Transl Med*. 2017. https://doi.org/10.4103/ctm.ctm_26_17.
91. Oronsky B, Carter C, Reid T, Brinkhaus F, Knox SJ. Just eat it: a review of CD47 and SIRP- α antagonism. *Semin Oncol*. 2020;47:117–24.

92. Petrova PS, Viller NN, Wong M, Pang X, Lin GHY, Dodge K, et al. TTI-621 (SIRPaFc): a CD47-blocking innate immune checkpoint inhibitor with broad antitumor activity and minimal erythrocyte binding. *Clin Cancer Res.* 2017;23:1068–79.
93. Brooke G, Holbrook JD, Brown MH, Barclay AN. Human lymphocytes interact directly with CD47 through a novel member of the signal regulatory protein (SIRP) family. *J Immunol.* 2004;173:2562–70.
94. Seiffert M, Brossart P, Cant C, Cella M, Colonna M, Brugger W, et al. Signal-regulatory protein alpha (SIRPalpha) but not SIRPbeta is involved in T-cell activation, binds to CD47 with high affinity, and is expressed on immature CD34(+)CD38(-) hematopoietic cells. *Blood.* 2001;97:2741–9.
95. Willingham SB, Volkmer J-P, Gentles AJ, Sahoo D, Dalerba P, Mitra SS, et al. The CD47-signal regulatory protein alpha (SIRPa) interaction is a therapeutic target for human solid tumors. *Proc Natl Acad Sci.* 2012. <https://doi.org/10.1073/pnas.1121623109>.
96. Yeh KC, Wu SH, Murphy JT, Lagarias JC. A cyanobacterial phytochrome two-component light sensory system. *Science.* 1997;277:1505–8.
97. Mörner CT. Untersuchung der proteinsubstanzen in den leichtbrechenden medien des auges I. *De Gruyter.* 1894;18:61–106.
98. de Jong WW, Leunissen JA, Voorter CE. Evolution of the alpha-crystallin/small heat-shock protein family. *Mol Biol Evol.* 1993;10:103–26.
99. Crandall KA, Hillis DM. Rhodopsin evolution in the dark. *Nature.* 1997;387:667–8.
100. Chang BSW, Jönsson K, Kazmi MA, Donoghue MJ, Sakmar TP. Recreating a functional ancestral archosaur visual pigment. *Mol Biol Evol.* 2002;19:1483–9.
101. Liu Y, Cui Y, Chi H, Xia Y, Liu H, Rossiter SJ, et al. Scotopic rod vision in tetrapods arose from multiple early adaptive shifts in the rate of retinal release. *Proc Natl Acad Sci USA.* 2019;116:12627–8.
102. Yokoyama S, Tada T, Zhang H, Britt L. Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proc Natl Acad Sci USA.* 2008;105:13480–5.
103. Pohl N, Sison-Mangus MP, Yee EN, Liswi SW, Briscoe AD. Impact of duplicate gene copies on phylogenetic analysis and divergence time estimates in butterflies. *BMC Evol Biol BioMed Central.* 2009;9:1–16.
104. Dornburg A, Santini F, Alfaro ME. The influence of model averaging on clade posteriors: an example using the triggerfishes (Family Balistidae). *Syst Biol.* 2008;57:905–19.
105. Dornburg A, Near TJ. The emerging phylogenetic perspective on the evolution of actinopterygian fishes. *Ann Rev Ecol Syst.* 2021. <https://doi.org/10.1146/annurev-ecolsys-122120-122554>.
106. Yu Z, Fischer R. Light sensing and responses in fungi. *Nat Rev Microbiol.* 2018;17:25–36.
107. Vierstra RD. Cyanophytochromes, bacteriophytochromes, and plant phytochromes. *Histidine Kinases Signal Transduct.* 2003. <https://doi.org/10.1016/b978-012372484-7/50014-x>.
108. Rodriguez-Romero J, Hedtke M, Kastner C, Müller S, Fischer R. Fungi, hidden in soil or up in the air: light makes a difference. *Ann Rev Microbiol.* 2010. <https://doi.org/10.1146/annurev.micro.112408.134000>.
109. Corrochano LM, Kuo A, Marcet-Houben M, Polaino S, Salamov A, Villalobos-Escobedo JM, et al. Expansion of signal transduction pathways in fungi by extensive genome duplication. *Curr Biol.* 2016;26:1577–84.
110. Corrochano LM. Light in the fungal world: from photoreception to gene transcription and beyond. *Annu Rev Genet.* 2019;53:149–70.
111. Wang Z, Wang J, Li N, Li J, Trail F, Dunlap JC, et al. Light sensing by opsins and fungal ecology: NOP-1 modulates entry into sexual reproduction in response to environmental cues. *Mol Ecol.* 2018;27:216–32.
112. Wang Z, Li N, Li J, Dunlap JC, Trail F, Townsend JP. The fast-evolving phy-2 gene modulates sexual development in response to light in the model fungus *Neurospora crassa*. *MBio.* 2016;7:e02148.
113. Wistow G. The human crystallin gene families. *Hum Genom BioMed Central.* 2012;6:1–10.
114. Wistow G, Slingsby C. Structure and evolution of crystallins. In: *Encyclopedia of the eye.* Academic Press; 2010. p. 229–38.
115. Kappé G, Purkiss AG, van Genesen ST, Slingsby C, Lubsen NH. Explosive expansion of betagamma-crystallin genes in the ancestral vertebrate. *J Mol Evol.* 2010;71:219–30.
116. Mackay DS, Andley UP, Shiels A. Cell death triggered by a novel mutation in the alphaA-crystallin gene underlies autosomal dominant cataract linked to chromosome 21q. *Eur J Hum Genet.* 2003;11:784–93.
117. Litt M, Kramer P, LaMorticella DM, Murphey W, Lovrien EW, Weleber RG. Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. *Hum Mol Genet.* 1998;7:471–4.
118. Devi RR, Yao W, Vijayalakshmi P, Sergeev YV, Sundaresan P, Fielding HJ. Crystallin gene mutations in Indian families with inherited pediatric cataract. *Mol Vis.* 2008;14:1157.
119. Brakenhoff RH, Aarts HJ, Reek FH, Lubsen NH, Schoenmakers JG. Human gamma-crystallin genes: a gene family on its way to extinction. *J Mol Biol.* 1990;216:519–32.
120. Lubsen NH, Aarts HJ, Schoenmakers JG. The evolution of lenticular proteins: the beta- and gamma-crystallin super gene family. *Prog Biophys Mol Biol.* 1988;51:47–76.
121. Ovchinnikov YuA. Rhodopsin and bacteriorhodopsin: structure-function relationships. *FEBS Lett.* 1982;148:179–91.
122. Nathans J, Hogness DS. Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell.* 1983;34:807–14.
123. Terakita A. The opsins. *Genome Biol BioMed Central.* 2005;6:1–9.
124. Chi H, Cui Y, Rossiter SJ, Liu Y. Convergent spectral shifts to blue-green vision in mammals extends the known sensitivity of vertebrate M/LWS pigments. *Proc Natl Acad Sci USA.* 2020;117:8303–5.
125. Nathans J, Thomas D, Hogness DS. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science.* 1986. <https://doi.org/10.1126/science.2937147>.
126. Musilova Z, Salzburger W, Cortesi F. The visual opsin gene repertoires of teleost fishes: evolution, ecology, and function. *Annu Rev Cell Dev Biol.* 2021;37:441–68.
127. Lin J-J, Wang F-Y, Li W-H, Wang T-Y. The rises and falls of opsin genes in 59 ray-finned fish genomes and their implications for environmental adaptation. *Sci Rep.* 2017;7:15568.
128. Zhao H, Rossiter SJ, Teeling EC, Li C, Cotton JA, Zhang S. The evolution of color vision in nocturnal mammals. *Proc Natl Acad Sci USA.* 2009;106:8980–5.
129. Eaton KM, Bernal MA, Backenstose NJC, Yule DL, Krabbenhoft TJ. Nanopore amplicon sequencing reveals molecular convergence and local adaptation of rhodopsin in great lakes salmonids. *Genom Biol Evol.* 2021. <https://doi.org/10.1093/gbe/evaa237>.
130. Hill J, Enbody ED, Pettersson ME, Sprehn CG, Bekkevold D, Folkvord A, et al. Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proc Natl Acad Sci USA.* 2019;116:18473–8.
131. Yoder EB, Parker CE, Tew A, Jones CD, Dornburg A. Decoupled spectral tuning and eye size diversification patterns in an Antarctic adaptive radiation. *bioRxiv.* 2022. <https://doi.org/10.1101/2022.09.28.509872>
132. Berry MH, Holt A, Salari A, Veit J, Visel M, Levitz J, et al. Restoration of high-sensitivity and adapting vision with a cone opsin. *Nat Commun.* 2019;10:1221.
133. Davidoff C. Cone opsin gene variants in color blindness and other vision disorders. 2015.
134. Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, et al. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc Natl Acad Sci USA.* 2009;106:13410–4.
135. Fernández R, Gabaldón T. Gene gain and loss across the metazoan tree of life. *Nat Ecol Evol.* 2020;4:524–33.
136. Suh S, Choi EH, Atanaskova MN. The expression of opsins in the human skin and its implications for photobiomodulation: a systematic review. *Photodermatol Photoimmunol Photomed.* 2020;36:329–38.
137. Moraes MN, de Assis LVM, Provencio I, de Castrucci AM. Opsins outside the eye and the skin: a more complex scenario than originally thought for a classical light sensor. *Cell Tissue Res.* 2021;385:519–38.
138. Mähnger LM, Roberts SB, Hanlon RT. Evidence for distributed light sensing in the skin of cuttlefish. *Sepia officinalis Biol Lett.* 2010;6:600–3.
139. Castellano-Pellicena I, Uzunbajakava NE, Mignon C, Raafs B, Botchkarev VA, Thornton MJ. Does blue light restore human epidermal barrier function via activation of Opsin during cutaneous wound healing? *Lasers Surg Med.* 2019;51:370–82.
140. Buck L, Axel R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell.* 1991;65:175–87.

141. Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell*. 1999;96:725–36.
142. Hildebrand JG, Shepherd GM. Mechanisms of olfactory discrimination: converging evidence for common principles across phyla. *Annu Rev Neurosci*. 1997;20:595–631.
143. Kaupp UB. Olfactory signalling in vertebrates and insects: differences and commonalities. *Nat Rev Neurosci*. 2010;11:188–200.
144. Hayden S, Bekaert M, Crider TA, Mariani S, Murphy WJ, Teeling EC. Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Res*. 2010;20:1–9.
145. Niimura Y. Olfactory receptor multigene family in vertebrates: from the viewpoint of evolutionary genomics. *Curr Genom*. 2012;13:103–14.
146. Niimura Y, Matsui A, Touhara K. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res*. 2014;24:1485–96.
147. Yohe LR, Fabbri M, Hanson M, Bhullar B-AS. Olfactory receptor gene evolution is unusually rapid across Tetrapoda and outpaces chemosensory phenotypic change. *Curr Zool*. 2020;66:505–14.
148. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. *Ann Rev Genet*. 2005. <https://doi.org/10.1146/annurev.genet.39.073003.112240>.
149. Niimura Y, Nei M. Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE*. 2007. <https://doi.org/10.1371/journal.pone.0000708>.
150. Sánchez-Gracia A, Vieira FG, Rozas J. Molecular evolution of the major chemosensory gene families in insects. *Heredity*. 2009. <https://doi.org/10.1038/hdy.2009.55>.
151. Bear DM, Lassance J-M, Hoekstra HE, Datta SR. The evolving neural and genetic architecture of vertebrate olfaction. *Curr Biol*. 2016;26:R1039–49.
152. Sato T, Hirono J, Hamana H, Ishikawa T, Shimizu A, Takashima I, et al. Architecture of odor information processing in the olfactory system. *Anat Sci Int*. 2008;83:195–206.
153. Dehara Y, Hashiguchi Y, Matsubara K, Yanai T, Kubo M, Kumazawa Y. Characterization of squamate olfactory receptor genes and their transcripts by the high-throughput sequencing approach. *Genome Biol Evol*. 2012;4:602–16.
154. McBride CS. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proc Natl Acad Sci USA*. 2007;104:4996–5001.
155. McBride CS, Arguello JR, O'Meara BC. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics*. 2007;177:1395–416.
156. Hayden S, Bekaert M, Goodbla A, Murphy WJ, Dávalos LM, Teeling EC. A cluster of olfactory receptor genes linked to frugivory in bats. *Mol Biol Evol*. 2014;31:917–27.
157. Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG, Whiteman NK. Evolution of herbivory in *Drosophilidae* linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proc Natl Acad Sci USA*. 2015;112:3026–31.
158. Gould F, Estock M, Hillier NK, Powell B, Groot AT, Ward CM, et al. Sexual isolation of male moths explained by a single pheromone response QTL containing four receptor genes. *Proc Natl Acad Sci USA*. 2010;107:8660–5.
159. Ferrero DM, Lemon JK, Fluegge D, Pashkovski SL, Korzan WJ, Datta SR, et al. Detection and avoidance of a carnivore odor by prey. *Proc Natl Acad Sci USA*. 2011;108:11235–40.
160. Hallem EA, Carlson JR. Coding of odors by a receptor repertoire. *Cell*. 2006;125:143–60.
161. Malnic B, Hirono J, Sato T, Buck LB. Combinatorial receptor codes for odors. *Cell*. 1999;96:713–23.
162. Magklara A, Lomvardas S. Stochastic gene expression in mammals: lessons from olfaction. *Trends Cell Biol*. 2013;23:449–56.
163. Nara K, Saraiva LR, Ye X, Buck LB. A large-scale analysis of odor coding in the olfactory epithelium. *J Neurosci*. 2011;31:9179–91.
164. Rodríguez I. Singular expression of olfactory receptor genes. *Cell*. 2013;155:274–7.
165. McClintock TS, Adipietro K, Titlow WB, Breheny P, Walz A, Mombaerts P, et al. In vivo identification of eugenol-responsive and muscone-responsive mouse odorant receptors. *J Neurosci*. 2014;34:15669–78.
166. Bushdid C, Magnasco MO, Vosshall LB, Keller A. Humans can discriminate more than 1 trillion olfactory stimuli. *Science*. 2014. <https://doi.org/10.1126/science.1249168>.
167. Haverkamp A, Hansson BS, Knaden M. Combinatorial codes and labeled lines: how insects use olfactory cues to find and judge food, mates, and oviposition sites in complex environments. *Front Physiol*. 2018;9:49.
168. Gemmell NJ, Rutherford K, Prost S, Tollis M, Winter D, Macey JR, et al. The tuatara genome reveals ancient features of amniote evolution. *Nature*. 2020;584:403–9.
169. Drum Z, Lanno S, Gregory S, Shimshak S, Barr W, Gatesman A, et al. Genomics analysis of *Drosophila sechellia* response to morinda citrifolia fruit diet. *G3*. 2022. <https://doi.org/10.1093/g3journal/jkac153>.
170. Shiao M-S, Chang J-M, Fan W-L, Lu M-YJ, Notredame C, Fang S, et al. Expression divergence of chemosensory genes between *Drosophila sechellia* and its sibling species and its implications for host shift. *Genome Biol Evol*. 2015;7:2843–58.
171. Drum ZA, Lanno SM, Gregory SM, Shimshak SJ, Ahamed M, Barr W, et al. Genomics analysis of hexanoic acid exposure in *Drosophila* species. *G3*. 2022. <https://doi.org/10.1093/g3journal/jkab354>.
172. Auer TO, Khallaf MA, Silbering AF, Zappia G, Ellis K, Álvarez-Ocaña R, et al. Olfactory receptor and circuit evolution promote host specialization. *Nature*. 2020;579:402–8.
173. Prieto-Godino LL, Rytz R, Cruchet S, Bargeton B, Abuin L, Silbering AF, et al. Evolution of acid-sensing olfactory circuits in *Drosophilids*. *Neuron*. 2017;93:661–76.e6.
174. Dressler RL. Biology of the orchid bees (Euglossini). *Ann Rev Ecol Syst*. 1982. <https://doi.org/10.1146/annurev.es.13.110182.002105>.
175. Ackerman JD. Specificity and mutual dependency of the orchid-euglossine bee interaction. *Biol J Linnean Soc*. 1983;20:301–14. <https://doi.org/10.1111/j.1095-8312.1983.tb01878.x>.
176. Cameron SA. Phylogeny and biology of neotropical orchid bees (Euglossini). *Annu Rev Entomol*. 2004;49:377–404.
177. Kimsey LS. The behaviour of male orchid bees (Apidae, Hymenoptera, Insecta) and the question of leks. *Anim Behav*. 1980;28(4):996–1004. [https://doi.org/10.1016/S0003-3472\(80\)80088-1](https://doi.org/10.1016/S0003-3472(80)80088-1).
178. Eltz T, Sager A, Lunau K. Juggling with volatiles: exposure of perfumes by displaying male orchid bees. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol*. 2005;191:575–81.
179. Zimmermann Y, Roubik DW, Eltz T. Species-specific attraction to pheromonal analogues in orchid bees. *Behav Ecol Sociobiol*. 2006. <https://doi.org/10.1007/s00265-006-0227-8>.
180. Pokorný T, Vogler I, Losch R, Schlütting P, Juárez P, Bissantz N, et al. Blown by the wind: the ecology of male courtship display behavior in orchid bees. *Ecology*. 2017;98:1140–52.
181. Stern DL, Dudley TR. Wing buzzing by male orchid bees, *Eulaema meriana* (Hymenoptera: Apidae). *J Kansas Entomol Soc*. 1991;64:88–94.
182. Dodson CH. Ethology of some bees of the tribe Euglossini (Hymenoptera: Apidae). *J Kansas Entomol Soc*. 1966;39:607–29.
183. Zimmermann Y, Ramírez SR, Eltz T. Chemical niche differentiation among sympatric species of orchid bees. *Ecology*. 2009;90:2994–3008.
184. Weber MG, Mitko L, Eltz T, Ramírez SR. Macroevolution of perfume signalling in orchid bees. *Ecol Lett*. 2016;19:1314–23.
185. Brand P, Ramírez SR, Leese F, Quezada-Euan JGG, Tollrian R, Eltz T. Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini). *BMC Evol Biol*. 2015;15:176.
186. Brand P, Saleh N, Pan H, Li C, Kapheim KM, Ramírez SR. The nuclear and mitochondrial genomes of the facultatively eusocial orchid bee. *G3*. 2017;7:2891–8.
187. Yoder AD, Larsen PA. The molecular evolutionary dynamics of the vomeronasal receptor (class 1) genes in primates: a gene family on the verge of a functional breakdown. *Front Neuroanat*. 2014;8:153.
188. Yohe LR. Ecological constraints on highly evolvable olfactory receptor genes and morphology in neotropical bats. *Evolution*. 2022. <https://doi.org/10.1111/evo.14591>.
189. Arguello JR, Roman Arguello J, Cardoso-Moreira M, Grenier JK, Gottipati S, Clark AG, et al. Extensive local adaptation within the chemosensory system following *Drosophila melanogaster*'s global expansion. *Nat Commun*. 2016. <https://doi.org/10.1038/ncomms11855>.

190. Yohe LR, Brand P. Handling editor: *Rebecca Fulle*: evolutionary ecology of chemosensation and its role in sensory drive. *Curr Zool.* 2018;64:525–33.
191. Moriya-Ito K, Hayakawa T, Suzuki H, Hagino-Yamagishi K, Nikaido M. Evolution of vomeronasal receptor 1 (V1R) genes in the common marmoset (*Callithrix jacchus*). *Gene.* 2018;642:343–53.
192. Perret M. Environmental and social determinants of sexual function in the male lesser mouse lemur (*Microcebus murinus*). *Folia Primatol.* 1992;59:1–25.
193. Aujard F. Effect of vomeronasal organ removal on male socio-sexual responses to female in a prosimian primate (*Microcebus murinus*). *Physiol Behav.* 1997. [https://doi.org/10.1016/s0031-9384\(97\)00206-0](https://doi.org/10.1016/s0031-9384(97)00206-0).
194. Buesching CD, Heistermann M, Hodges JK, Zimmermann E. Multimodal oestrus advertisement in a small nocturnal prosimian, *Microcebus murinus*. *Folia Primatol.* 1998. <https://doi.org/10.1159/000052718>.
195. Eberle M, Kappeler PM. Sex in the dark: determinants and consequences of mixed male mating tactics in *Microcebus murinus*, a small solitary nocturnal primate. *Behav Ecol Sociobiol.* 2004;57(1):77–90. <https://doi.org/10.1007/s00265-004-0826-1>.
196. Wynn EH, Sánchez-Andrade G, Carss KJ, Logan DW. Genomic variation in the vomeronasal receptor gene repertoires of inbred mice. *BMC Genomics.* 2012;13:415.
197. Grus WE, Zhang J. Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats. *Gene.* 2004;340:303–12.
198. Lane RP, Young J, Newman T, Trask BJ. Species specificity in rodent pheromone receptor repertoires. *Genome Res.* 2004;14:603–8.
199. Park SH, Podlaha O, Grus WE, Zhang J. The microevolution of V1r vomeronasal receptor genes in mice. *Genome Biol Evol.* 2011;3:401–12.
200. Herrera JP. Testing the adaptive radiation hypothesis for the lemurs of Madagascar. *R Soc Open Sci.* 2017;4:161014.
201. Herrera JP, Dávalos LM. Phylogeny and divergence times of lemurs inferred with recent and ancient fossils in the tree. *Syst Biol.* 2016;65:772–91.
202. Yohe LR, Davies KTJ, Rossiter SJ, Dávalos LM. Expressed vomeronasal type-1 receptors (V1rs) in bats uncover conserved sequences underlying social chemical signaling. *Genome Biol Evol.* 2019;11:2741–9.
203. Adipietro KA, Mainland JD, Matsunami H. Functional evolution of mammalian odorant receptors. *PLoS Genet.* 2012;8(7):e1002821. <https://doi.org/10.1371/journal.pgen.1002821>.
204. Han W, Yiran W, Zeng L, Zhao S. Building the chordata olfactory receptor database using more than 400,000 receptors annotated by genome2or. *Sci China Life Sci.* 2022. <https://doi.org/10.1007/s11427-021-2081-6>.
205. Gonzalez FJ, Nebert DW. Evolution of the P450 gene superfamily: animal-plant “warfare”, molecular drive and human genetic differences in drug oxidation. *Trends Genet.* 1990;6:182–6.
206. Nelson DR, Zeldin DC, Hoffman SMG, Maltais LJ, Wain HM, Nebert DW. Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics.* 2004;14:1–18.
207. Nebert DW. Aryl hydrocarbon receptor (AHR): “pioneer member” of the basic-helix/loop/helix per-Arnt-sim (bHLH/PAS) family of “sensors” of foreign and endogenous signals. *Prog Lipid Res.* 2017;67:38–57.
208. Davenport ER, Sanders JG, Song SJ, Amato KR, Clark AG, Knight R. The human microbiome in evolution. *BMC Biol.* 2017. <https://doi.org/10.1186/s12915-017-0454-7>.
209. Schwertmann L, Focke O, Dirks J-H. Morphology, shape variation and movement of skeletal elements in starfish (*Asterias rubens*). *J Anat.* 2019;234:656–67.
210. Nebert DW. Proposed role of drug-metabolizing enzymes: regulation of steady state levels of the ligands that effect growth, homeostasis, differentiation, and neuroendocrine functions. *Mol Endocrinol.* 1991;5:1203–14.
211. Nebert DW. Drug-metabolizing enzymes in ligand-modulated transcription. *Biochem Pharmacol.* 1994;47:25–37.
212. Pascussi J-M, Gerbal-Chaloin S, Duret C, Daujat-Chavanieu M, Vilarem M-J, Maurel P. The tangle of nuclear receptors that controls xenobiotic metabolism and transport: crosstalk and consequences. *Annu Rev Pharmacol Toxicol.* 2008;48:1–32.
213. Nebert DW, Wikvall K, Miller WL. Human cytochromes P450 in health and disease. *Philos Trans R Soc Lond B Biol Sci.* 2013;368:20120431.
214. Scheer N, Kapelyukh Y, Chatham L, Rode A, Buechel S, Wolf CR. Generation and characterization of novel cytochrome P450 Cyp2c gene cluster knockout and CYP2C9 humanized mouse lines. *Mol Pharmacol.* 2012;82:1022–9.
215. Shows TB, Alper CA, Bootsma D, Dorf M, Douglas T, Huisman T, et al. International system for human gene nomenclature (1979) ISGN (1979). *Cytogenet Cell Genet.* 1979;25:96–116.
216. Shows TB, McAlpine PJ, Boucheix C, Collins FS, Conneally PM, Frézal J, et al. Guidelines for human gene nomenclature: an international system for human gene nomenclature (ISGN, 1987). *Cytogenet Cell Genet.* 1987;46:11–28.
217. Bruford EA, Braschi B, Denny P, Jones TEM, Seal RL, Tweedie S. Guidelines for human gene nomenclature. *Nat Genet.* 2020;52:754–8.
218. Snell GD. Gene and chromosome mutations. In: Little CC, Snell GD, editors. *Biology of the laboratory mouse*. Philadelphia: Blakiston Co.; 2012. p. 34–247.
219. Borrego F. The CD300 molecules: an emerging family of regulators of the immune system. *Blood.* 2013;121:1951–60.
220. Vitallé J, Terrén I, Orrantia A, Zenarruzabeitia O, Borrego F. CD300 receptor family in viral infections. *Eur J Immunol.* 2019;49:364–74.
221. Vitallé J, Terrén I, Orrantia A, Bilbao A, Gamboa PM, Borrego F, et al. The expression and function of CD300 molecules in the main players of allergic responses: mast cells, basophils and eosinophils. *Int J Mol Sci.* 2020. <https://doi.org/10.3390/ijms21093173>.
222. Nebert DW, Gonzalez FJ. P450 genes: structure, evolution, and regulation. *Annu Rev Biochem.* 1987;56:945–93.
223. Nebert DW, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ, Guengerich FP, et al. The P450 gene superfamily: recommended nomenclature. *DNA.* 1987;6:1–11.
224. Nebert DW, Nelson DR, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ, et al. The P450 superfamily: updated listing of all genes and recommended nomenclature for the chromosomal loci. *DNA.* 1989;8:1–13.
225. Hansen CC, Nelson DR, Møller BL, Werck-Reichhart D. Plant cytochrome P450 plasticity and evolution. *Mol Plant.* 2021;14:1244–65.
226. Agnarsson I, Kuntner M. Taxonomy in a changing world: seeking solutions for a science in crisis. *Syst Biol.* 2007. <https://doi.org/10.1080/10635150701424546>.
227. Olender T, Jones TEM, Bruford E, Lancet D. A unified nomenclature for vertebrate olfactory receptors. *BMC Evol Biol.* 2020;20:42.
228. Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, et al. Gene-names.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res.* 2021;49(D1):D939–46.
229. McCarthy FM, Jones TEM, Kwitek AE, Smith CL, Vize PD, Westerfield M, et al. The case for standardising gene nomenclature across vertebrates. *Preprints; 2021* [cited 2022 Sep 8]; <https://www.preprints.org/manuscript/202109.0485/v1>
230. Dornburg A, Ota T, Criscitiello MF, Irene Salinas J, Sunyer O, Magadán S, et al. From IgZ to IgT: a call for a common nomenclature for immunoglobulin heavy chain genes of ray-finned fish. *Zebrafish.* 2021;18:343–5. <https://doi.org/10.1089/zeb.2021.0071>.
231. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10.
232. Johnson M, Zaretskaya I, Raytselis Y, Merezhuik Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008;36:W5–9.
233. Jones DT, Swindells MB. Getting the most from PSI–BLAST. *Trends Biochem Sci Elsevier.* 2002;27:161–4.
234. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.
235. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2014;12:59–60.
236. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol.* 2011;7:e1002195.
237. Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 2009;23:205–11.
238. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019;47:D506–15.
239. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient

- methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37:1530–4.
240. Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, et al. RevBayes: bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst Biol Oxford Academic.* 2016;65:726–36.
 241. Lewis PO, Chen M-H, Kuo L, Lewis LA, Fučíková K, Neupane S, et al. Estimating bayesian phylogenetic information content. *Syst Biol.* 2016. <https://doi.org/10.1093/sysbio/syw042>.
 242. Salichos L, Leonidas S, Antonis R. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature.* 2013;497:327–31.
 243. Chen M-Y, Liang D, Zhang P. Selecting question-specific genes to reduce incongruence in phylogenomics: a case study of jawed vertebrate backbone phylogeny. *Syst Biol.* 2015;64:1104–20.
 244. Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJP. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. *Mol Biol Evol.* 2013;30:2134–44.
 245. Shen X-X, Hittinger CT, Rokas A. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat Ecol Evol.* 2017;1:126.
 246. Townsend JP, Su Z, Tekle YI. Phylogenetic signal and noise: predicting the power of a data set to resolve phylogeny. *Syst Biol.* 2012;61:835–49.
 247. Gilbert PS, Chang J, Pan C, Sobel EM, Sinsheimer JS, Faircloth BC, et al. Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. *Mol Phylogenet Evol.* 2015;92:140–6.
 248. Granados Mendoza C, Naumann J, Samain M-S, Goetghebeur P, De Smet Y, Wanke S. A genome-scale mining strategy for recovering novel rapidly-evolving nuclear single-copy genes for addressing shallow-scale phylogenetics in *Hydrangea*. *BMC Evol Biol.* 2015;15:132.
 249. Dornburg A, Townsend JP, Wang Z. Maximizing power in phylogenetics and phylogenomics: a perspective illuminated by fungal big data. *Adv Genet.* 2017;100:1–47.
 250. Dornburg A, Su Z, Townsend JP. Optimal rates for phylogenetic inference and experimental design in the era of genome-scale data sets. *Syst Biol.* 2019;68:145–56.
 251. Weisman CM, Murray AW, Eddy SR. Many, but not all, lineage-specific genes can be explained by homology detection failure. *PLoS Biol.* 2020;18:e3000862.
 252. Graybeal A. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. *Syst Biol.* 1994;43:174–93.
 253. Roje DM. Incorporating molecular phylogenetics with larval morphology while mitigating the effects of substitution saturation on phylogeny estimation: a new hypothesis of relationships for the flatfish family pleuronectidae (Percomorpha: Pleuronectiformes). *Mol Phylogenet Evol.* 2010;56:586–600.
 254. Mueller RL. Evolutionary rates, divergence dates, and the performance of mitochondrial genes in Bayesian phylogenetic analysis. *Syst Biol.* 2006;55:289–300.
 255. Dornburg A, Townsend JP, Brooks W, Spriggs E, Eytan RI, Moore JA, et al. New insights on the sister lineage of percomorph fishes with an anchored hybrid enrichment dataset. *Mol Phylogenet Evol.* 2017;110:27–38.
 256. Duchêne DA, Mather N, Van Der Wal C, Ho SYW. Excluding loci with substitution saturation improves inferences from phylogenomic data. *Syst Biol.* 2022;71:676–89.
 257. Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, et al. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 2011;9:e1000602.
 258. Field DJ, Berv JS, Hsiang AY, Lanfear R, Landis MJ, Dornburg A. Timing the extant avian radiation: The rise of modern birds, and the importance of modeling molecular rate variation. <https://doi.org/10.7287/peerj.preprints.27521>.
 259. Rosenfeld JA, DeSalle R. E value cutoff and eukaryotic genome content phylogenetics. *Mol Phylogenet Evol.* 2012;63(2):342–50. <https://doi.org/10.1016/j.ympev.2012.01.003>.
 260. Townsend JP, Lopez-Giraldez F. Optimal selection of gene and ingroup taxon sampling for resolving phylogenetic relationships. *Syst Biol.* 2010;59:446–57.
 261. Townsend JP, Leuenberger C. Taxon sampling and the optimal rates of evolution for phylogenetic inference. *Syst Biol.* 2011;60(3):358–65. <https://doi.org/10.1093/sysbio/syq097>.
 262. Betancur-R R, Li C, Munroe TA, Ballesteros JA, Ortí G. Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). *Syst Biol.* 2013;62:763–85.
 263. Lartillot N. Phylogenetic patterns of GC-biased gene conversion in placental mammals and the evolutionary dynamics of recombination landscapes. *Mol Biol Evol.* 2013;30:489–502.
 264. Townsend JP, López-Giraldez F, Friedman R. The phylogenetic informativeness of nucleotide and amino acid sequences for reconstructing the vertebrate tree. *J Mol Evol.* 2008;67:437–47.
 265. Dornburg A, Townsend JP, Friedman M, Near TJ. Phylogenetic informativeness reconciles ray-finned fish molecular divergence times. *BMC Evol Biol.* 2014;14:169.
 266. Parker E, Dornburg A, Domínguez-Domínguez O, Piller KR. Assessing phylogenetic information to reveal uncertainty in historical data: An example using Goodeinae (Teleostei: Cyprinodontiformes: Goodeidae). *Mol Phylogenet Evol.* 2019;134:282–90.
 267. Dornburg A, Fisk JN, Tamagnan J, Townsend JP. PhylInformR: phylogenetic experimental design and phylogenomic data exploration in R. *BMC Evol Biol.* 2016;16:262.
 268. Papp B, Pál C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. *Nature.* 2003;424:194–7.
 269. Kuraku S, Meyer A. Whole genome duplications and the radiation of vertebrates. In: Dittmar K, Liberles D, editors. *Evolution after gene duplication*. Hoboken: Wiley; 2010. p. 299–311.
 270. Ohno S. *Evolution by Gene Duplication*. Cham: Springer; 2014.
 271. Yokoyama S, Takenaka N. The molecular basis of adaptive evolution of squirrelfish rhodopsins. *Mol Biol Evol.* 2004;21:2071–8.
 272. Srouf JT, Losos JB. Ecological opportunity and adaptive radiation. *Ann Rev Ecol Evol Syst.* 2016. <https://doi.org/10.1146/annurev-ecolsys-121415-032254>.
 273. Dornburg A, Sidlauskas B, Santini F, Sorenson L, Near TJ, Alfaro ME. The influence of an innovative locomotor strategy on the phenotypic diversification of triggerfish (family: Balistidae). *Evolution.* 2011;65:1912–26.
 274. Price SA, Schmitz L, Oufiero CE, Eytan RI, Dornburg A, Smith WL, et al. Two waves of colonization straddling the K-Pg boundary formed the modern reef fish fauna. *Proc Biol Sci.* 2014;281:20140321.
 275. Daane JM, Dornburg A, Smits P, MacGuigan DJ, Brent Hawkins M, Near TJ, et al. Historical contingency shapes adaptive radiation in Antarctic fishes. *Nat Ecol Evol.* 2019;3:1102–9.
 276. Brawand D, Wagner CE, Li Yi, Malinsky M, Keller I, Fan S, et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature.* 2014;513:375–81.
 277. Gould SJ. *The structure of evolutionary theory*. Harvard: Harvard University Press; 2002.
 278. Rudnicki R, Tiuryn J, Wójciewicz D. A model for the evolution of paralog families in genomes. *J Math Biol Springer.* 2006;53:759–70.
 279. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics.* 2006;22(10):1269–71. <https://doi.org/10.1093/bioinformatics/btl1097>.
 280. Chauve C, Doyon J-P, El-Mabrouk N. Gene family evolution by duplication, speciation, and loss. *J Comput Biol.* 2008;15:1043–62. <https://doi.org/10.1089/cmb.2008.0054>.
 281. Abi-Rached L, Moesta AK, Rajalingam R, Guethlein LA, Parham P. Human-specific evolution and adaptation led to major qualitative differences in the variable receptors of human and chimpanzee natural killer cells. *PLoS Genet.* 2010;6:e1001192.
 282. Guethlein LA, Norman PJ, Heijmans CMC, de Groot NG, Hilton HG, Babrzadeh F, et al. Two orangutan species have evolved different KIR alleles and haplotypes. *J Immunol.* 2017;198:3157–69.
 283. Wroblewski EE, Parham P, Guethlein LA. Two to Tango: co-evolution of hominid natural killer cell receptors and MHC. *Front Immunol.* 2019;10:177.
 284. Mager DL, McQueen KL, Wee V, Freeman JD. Evolution of natural killer cell receptors: coexistence of functional Ly49 and KIR genes in baboons. *Curr Biol.* 2001;11:626–30.

285. Bruijnesteijn J, de Groot N, de Vos-Rouweler AJM, de Groot NG, Bontrop RE. Comparative genetics of KIR haplotype diversity in humans and rhesus macaques: the balancing act. *Immunogenetics*. 2022;74:313–26.
286. Cadavid LF, Lun C-M. Lineage-specific diversification of killer cell Ig-like receptors in the owl monkey, a New World primate. *Immunogenetics*. 2009;61:27–41.
287. Averdam A, Petersen B, Rosner C, Neff J, Roos C, Eberle M, et al. A novel system of polymorphic and diverse NK cell receptors in primates. *PLoS Genet*. 2009;5:e1000688.
288. Hoelsbrekken SE, Nylenna Ø, Saether PC, Slettedal IO, Ryan JC, Fossum S, et al. Cutting edge: molecular cloning of a killer cell Ig-like receptor in the mouse and rat. *J Immunol*. 2003;170:2259–63.
289. Sambrook JG, Sehra H, Coggill P, Humphray S, Palmer S, Sims S, et al. Identification of a single killer immunoglobulin-like receptor (KIR) gene in the porcine leukocyte receptor complex on chromosome 6q. *Immunogenetics*. 2006;58:481–6.
290. Barten R, Trowsdale J. The human Ly-49L gene. *Immunogenetics*. 1999;49:731–4.
291. Guethlein LA, Flodin LR, Adams EJ, Parham P. NK cell receptors of the orangutan (*Pongo pygmaeus*): a pivotal species for tracking the coevolution of killer cell Ig-like receptors with MHC-C. *J Immunol*. 2002;169:220–9.
292. Gagnier L, Wilhelm BT, Mager DL. Ly49 genes in non-rodent mammals. *Immunogenetics*. 2003;55:109–15.
293. Schwartz JC, Gibson MS, Heimeier D, Koren S, Phillippy AM, Bickhart DM, et al. The evolution of the natural killer complex; a comparison between mammals using new high-quality genome assemblies and targeted annotation. *Immunogenetics*. 2017;69:255–69.
294. Futas J, Oppelt J, Janova E, Musilova P, Horin P. Complex variation in the KLRA (LY49) immunity-related genomic region in horses. *Hladnikia*. 2020;96:257–67.
295. Holland HL, Weber HK. Enzymatic hydroxylation reactions. *Curr Opin Biotechnol*. 2000;11:547–53.
296. Bell EL, Finnigan W, France SP, Green AP, Hayes MA, Hepworth LJ, et al. *Biotocatalysis*. Nat Rev Methods Primers. 2021. <https://doi.org/10.1038/s43586-021-00044-z>.
297. Nelson DR. Cytochrome P450 and the individuality of species. *Arch Biochem Biophys*. 1999;369:1–10.
298. Hernandez D, Janmohamed A, Chandan P, Phillips IR, Shephard EA. Organization and evolution of the flavin-containing monooxygenase genes of human and mouse: identification of novel gene and pseudogene clusters. *Pharmacogenetics*. 2004;14:117–30.
299. Krueger SK, Williams DE. Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol Ther*. 2005;106:357–87.
300. Jörnvall H. MDR-alcohol dehydrogenases. *Chem Biol Interact*. 2017;276:75–6.
301. Holmes RS. Alcohol dehydrogenases: a family of isozymes with differential functions. *Alcohol Alcohol Suppl*. 1994;2:127–30.
302. Vasiliou V, Bairoch A, Tipton KF, Nebert DW. Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. *Pharmacogenetics*. 1999;9:421–34.
303. Shortall K, Djeghader A, Magner E, Soulimane T. Insights into aldehyde dehydrogenase enzymes: a structural perspective. *Front Mol Biosci*. 2021;8:659550.
304. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. *Exp Mol Med*. 2017;49:e324.
305. Black JC, Van Rechem C, Whetstone JR. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell*. 2012;48:491–507.
306. Edmondson DE, Binda C. Monoamine oxidases. *Subcell Biochem*. 2018;87:117–39.
307. Benedetti MS. Biotransformation of xenobiotics by amine oxidases. *Fundam Clin Pharmacol*. 2001;15:75–84.
308. de Oliveira FK, Santos LO, Buffon JG. Mechanism of action, sources, and application of peroxidases. *Food Res Int*. 2021;143:110266.
309. O'Brien PJ. Peroxidases. *Chem Biol Interact*. 2000;129:113–39.
310. Goyal MM, Basak A. Human catalase: looking for complete identity. *Protein Cell*. 2010;1:888–97.
311. Zamocky M, Furtmüller PG, Obinger C. Evolution of catalases from bacteria to humans. *Antioxid Redox Signal*. 2008;10:1527–48.
312. Martínez AT, Ruiz-Dueñas FJ, Camarero S, Serrano A, Linde D, Lund H, et al. Oxidoreductases on their way to industrial biotransformations. *Biotechnol Adv*. 2017;35:815–31.
313. Rendic S, Guengerich FP. Survey of human oxidoreductases and cytochrome p450 enzymes involved in the metabolism of xenobiotic and natural chemicals. *Chem Res Toxicol*. 2015;28:38–42.
314. Waskell L, Kim J-JP. Electron transfer partners of cytochrome P450. In: Paul R, de Montellano O, editors. *Cytochrome P450*. Cham: Springer; 2015. p. 33–68.
315. Chen S, Wu K, Knox R. Structure-function studies of DT-diaphorase (NQO1) and NRH: quinone oxidoreductase (NQO2). *Free Radic Biol Med*. 2000;29:276–84.
316. Penning TM. The aldo-keto reductases (AKRs): overview. *Chem Biol Interact*. 2015;234:236–46.
317. Forrest GL, Gonzalez B. Carbonyl reductase. *Chem Biol Interact*. 2000;129:21–40.
318. Kallberg Y, Oppermann U, Jörnvall H, Persson B. Short-chain dehydrogenase/reductase (SDR) relationships: a large family with eight clusters common to human, animal, and plant genomes. *Protein Sci*. 2002;11:636–41.
319. Dong J, Fernández-Fueyo E, Hollmann F, Paul CE, Pesic M, Schmidt S, et al. Biocatalytic oxidation reactions: a chemist's perspective. *Angew Chem Int Ed Engl*. 2018;57:9238–61.
320. Kodani SD, Hammock BD. The 2014 Bernard B. brodie award lecture—epoxide hydrolases: drug metabolism to therapeutics for chronic pain. *Drug Metab Dispos*. 2015;43(5):788–802. <https://doi.org/10.1124/dmd.115.063339>.
321. Gautheron J, Jéru I. The multifaceted role of epoxide hydrolases in human health and disease. *Int J Mol Sci*. 2020. <https://doi.org/10.3390/ijms22010013>.
322. Wu Z, Liu C, Zhang Z, Zheng R, Zheng Y. Amidase as a versatile tool in amide-bond cleavage: From molecular features to biotechnological applications. *Biotechnol Adv*. 2020;43: 107574.
323. Anthonen HW, Baptista A, Drabløs F, Martel P, Petersen SB, Sebastião M, et al. Lipases and esterases: a review of their sequences, structure and evolution. *Biotechnol Annu Rev*. 1995;1:315–71.
324. Fojan P, Jonson PH, Petersen MT, Petersen SB. What distinguishes an esterase from a lipase: a novel structural approach. *Biochimie*. 2000;82:1033–41.
325. Zechner R, Zimmermann R, Eichmann TO, Kohlwein SD, Haemmerle G, Lass A, et al. FAT SIGNALS—lipases and lipolysis in lipid metabolism and signaling. *Cell Metab*. 2012;15:279–91.
326. Meech R, Hu DG, McKinnon RA, Mubarak SN, Haines AZ, Nair PC, et al. The UDP-glycosyltransferase (UGT) superfamily: new members, new functions, and novel paradigms. *Physiol Rev*. 2019;99:1153–222.
327. Oda S, Fukami T, Yokoi T, Nakajima M. A comprehensive review of UDP-glucuronosyltransferase and esterases for drug development. *Drug Metab Pharmacokinet*. 2015;30:30–51.
328. Nebert DW, Vasiliou V. Analysis of the glutathione S-transferase (GST) gene family. *Hum Genom*. 2004;1:460–4.
329. Blanchard RL, Freimuth RR, Buck J, Weinsilboum RM, Coughtrie MWH. A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily. *Pharmacogenetics*. 2004;14:199–211.
330. Pedersen LC, Yi M, Pedersen LG, Kaminski AM. From steroid and drug metabolism to glycobiology, using sulfotransferase structures to understand and tailor function. *Drug Metab Dispos*. 2022;50:1027–41.
331. Vetting MW, de Carvalho LPS, Yu M, Hegde SS, Magnet S, Roderick SL, et al. Structure and functions of the GNAT superfamily of acetyltransferases. *Arch Biochem Biophys*. 2005;433:212–26.
332. Sim E, Fakis G, Laurieri N, Boukouvala S. Arylamine N-acetyltransferases—from drug metabolism and pharmacogenetics to identification of novel targets for pharmacological intervention. *Adv Pharmacol*. 2012;63:169–205.
333. Divanovic S, Dalli J, Jorge-Nebert LF, Flick LM, Gálvez-Peralta M, Boespflug ND, et al. Contributions of the three CYP1 monooxygenases to pro-inflammatory and inflammation-resolution lipid mediator pathways. *J Immunol*. 2013;191:3347–57.

334. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther.* 2013;138:103–41.
335. Brand P, Hinojosa-Díaz IA, Ayala R, Daigle M, Yurrita Obiols CL, Eltz T, et al. The evolution of sexual signaling is linked to odorant receptor tuning in perfume-collecting orchid bees. *Nat Commun.* 2020;11:244.
336. Eltz T, Zimmermann Y, Pfeiffer C, Pech JR, Twele R, Francke W, et al. An olfactory shift is associated with male perfume differentiation and species divergence in orchid bees. *Curr Biol.* 2008;18:1844–8.
337. Ramírez SR, Eltz T, Fujiwara MK, Gerlach G, Goldman-Huertas B, Tsutsui ND, et al. Asynchronous diversification in a specialized plant-pollinator mutualism. *Science.* 2011;333:1742–6.
338. Eltz T, Fritzsche F, Pech JR, Zimmermann Y, Ramírez SR, Quezada-Euan JG, et al. Characterization of the orchid bee *Euglossa viridissima* (Apidae: Euglossini) and a novel cryptic sibling species, by morphological, chemical, and genetic characters. *Zool J Linnean Soc.* 2011;163:1064–76. <https://doi.org/10.1111/j.1096-3642.2011.00740.x>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

