Mouse protein coding diversity: What’s left to discover?

Jingtao Lilue1,2, Anu Shivalikanjli1, David J. Adams3, Thomas M. Keane1,4*

1 European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, United Kingdom, 2 Instituto Gulbenkian de Ciência, Oeiras, Lisbon, Portugal, 3 Wellcome Sanger Institute, Hinxton, United Kingdom, 4 School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

* tk2@ebi.ac.uk

Abstract

For over a century, mice have been used to model human disease, leading to many fundamental discoveries about mammalian biology and the development of new therapies. Mouse genetics research has been further catalysed by a plethora of genomic resources developed in the last 20 years, including the genome sequence of C57BL/6J and more recently the first draft reference genomes for 16 additional laboratory strains. Collectively, the comparison of these genomes highlights the extreme diversity that exists at loci associated with the immune system, pathogen response, and key sensory functions, which form the foundation for dissecting phenotypic traits in vivo. We review the current status of the mouse genome across the diversity of the mouse lineage and discuss the value of mice to understanding human disease.

Introduction

Although mice and humans have coexisted for many millennia, modern mouse genetics was initiated in the early 20th century [1]. The first genetically homozygous mouse strain was DBA [2], developed to study coat color inheritance and cancer susceptibility. Subsequently, hundreds of genetically defined strains for modelling human diseases and biological processes (e.g., behaviour, carcinogenesis, and immune response against pathogens) were developed. As one of the most important model organisms in biomedical research, the mouse was the second
mammalian genome to be sequenced after the human genome [3]. The C57BL/6J reference genome has enabled the creation of detailed molecular maps of mouse diversity [4,5], generation of null alleles, and phenotyping across thousands of genes; and enabled genetic screens at an unprecedented rate [6].

Modern-day laboratory mouse strains are comprised of classical and wild-derived strains. The classical inbred strains have ‘fancy mice’ as their founding ancestors and are largely Mus musculus domesticus derived. Other subspecies—M. musculus musculus and M. musculus castaneus—contribute approximately 4%–14% to the classical strains [7]. A genome-wide haplotype map of 100 classical mouse strains showed that over 97% of their genome is a mosaic of less than 10 haplotypes [8]. Nevertheless, there are many loci in classical mouse strains, like the major histocompatibility complex (MHC), which have extensive haplotypic diversity [9,10].

Wild-derived inbred mouse strains are recent progenies of wild-caught individuals of M. musculus musculus, M. musculus castaneus, and M. spretus origin and therefore contain many divergent haplotypes not shared with classical inbred strains. Wild-derived strains are increasingly employed as mouse models to study phenotypes—such as resistance against Orthomyxovirus [11], virulent Toxoplasma gondii strains (CIM, CAST/EiJ, and PWK/PhJ) [12,13], resistance to anticoagulant rodenticides (SPRET/EiJ) [14], and resistance to cerebral Plasmodium berghei (WLA/Pas) [15].

**Limitations of a single mouse reference genome**

Wild-derived mouse strains have hundreds of thousands of structural differences and novel haplotypes compared to C57BL/6J [4,5]. Most SNP discoveries, if not all, are based on high-density genotyping or short-read sequencing. Paired-end reads are aligned to the C57BL/6J reference genome to identify SNPs, indels, and structure variations (SVs) [16,17]. This means that using the C57BL/6J reference genome to study these strains is blind to many nonreference loci [16]. In these strain-specific diverse regions (SSDRs), next-generation sequencing (NGS) reads are forced to map incorrectly to other paralogous loci in the reference and are often represented as dense regions of heterozygous SNPs (hSNPs) that disrupt the collinearity between the genome of a mouse strain and the reference [13,16]. SSDRs are enriched for genes associated with immunity, sensory, sexual reproduction, and behaviour [16]. In this review, we will introduce the SSDRs among 16 mouse strains and their potential importance in human biomedical research.

Individual SSDRs associated with phenotypes in mouse inbred strains have long been studied (Fig 1). In 2004, a high resolution whole genome Bacterial Artificial Chromosome (BAC) array analysis reported ‘segmental polymorphisms’ between mouse strains C57BL/6 and 129/Sv [18]. Subsequent work found similar patterns by comparative genomic hybridization analysis and reported 2,094 ‘copy number variations’ (CNVs) in 41 inbred strains [19]. In 2015, an analysis of 351 high-density microarray data for mouse tail samples highlighted 9,634 putative autosomal CNVs affecting 6.87% of the mouse genome [20]. In 2016, Morgan and colleagues performed a genome-wide subsequence diversity test in seven mouse strains and two wild mice samples and reported at least 0.8% of the mouse genome is a ‘genomic revolving door’ with high mutation and recombination rates [21]. In 2018, the first draft de novo assemblies of 16 mouse strains successfully assembled some of these regions, reporting a total of 2,567 SSDRs that encompass 0.5%–2.8% of the mouse genome (Fig 2A and S1 Data), encoding 1,828 coding genes. These genes can be classified into 468 gene families (S1 Table), and 318 (67.7%) have previously been studied in detail. Only 3.1% of gene families have complete sequences (introns and intergenic regions) for multiple mouse strains, 9.8% have coding regions from multiple mouse strains, and most (87.1%) studies draw scientific conclusions based on a single
laboratory mouse strain, typically C57BL/6J or a 129 substrain (Fig 2B). SSDRs are enriched for recently transposed long interspersed nuclear elements (LINEs) and long-terminal repeat (LTR) elements, posing a challenge for genome assembly [16] and consequently are often incomplete in the current mouse reference genome.

Immune-related genes in SSDRs

SSDRs are highly enriched for immunity and infection response related genes (Fig 2C). Examples include MHC, natural killer gene complex [22,23], T-cell receptors [24], and immunoglobulin variable regions [25], which play central roles in non-self-recognition and adaptive immunity. Other loci include oligoadenylate-synthetase 1 (Oas1) complex [26], AIM2-like receptors [27], and Schlafen gene family [28] for virus innate immune response; NOD-like receptors—Nlrp1—for anthrax lethal toxin resistance [29]; immunity-related GTPases (IRGs) for intracellular pathogen resistance [13]; and α and β defensins for immunomodulatory and antimicrobial function in intestinal crypts [30]. An interesting example is the Intelectin (Itln) members encoded on chromosome 1. Intelectin is known to be highly up-regulated in the immune response to parasitic infections, e.g., Trichinella spiralis. In the C57BL/6J reference genome, only one Itln allele can be found; however, BALB/cJ has two (Itln1 and Itlnb) [31], and strain 129S7 has up to 6 Itln alleles (Itln1 to Itln6) [32]. Similarly, the adjacent gene Natural Killer Cell Receptor 2B4 (Cd244) shows similar patterns of CNVs in recent de novo assemblies [16].

Recent de novo assemblies have also highlighted many loci with polymorphisms previously unreported in mice. Apolipoprotein L (APOL) members encoded on chromosome 15 show high levels of CNVs in classical and wild-derived mouse strains [16]. There are very few studies in murine Apol members; however, their orthologues in humans (APOL1) are polymorphic [33]. Some alleles confer resistance to Trypanosoma brucei brucei in humans but at the same time lead to chronic kidney disease [33]. Skint gene family members, named after 'skin
thickness' because they regulate epidermal γδ T cells, are associated with chronic wound healing deficiencies in humans [34,35]. A single SNP was reported in mouse strain FVB Tac, which causes selective deficiency for epidermal Vγ5+Vδ1+ T cells [35]. However, the polymorphism of the Skint family appears to be much more complex than previously reported among mouse strains. Eosinophil-Associated RNases (Ears) encoded on chromosome 14 are orthologues of human eosinophil-derived neurotoxin (EDN) and eosinophil cationic protein (ECP), which are highly charged cytotoxic proteins released from activated eosinophil granules [34]. Mouse Ears can promote virus clearance [36] and play a role in the Schistosoma resistance [37]. Although evidence of positive selection has been found for Ears members in the reference genome, their diversity between inbred strains is poorly documented. At least three haplotypes can be found in classical inbred mouse strains (haplotype1: C57BL/6J, C57BL/6NJ, 129S1, AKR/J, BALB/cJ, A/J, CBA/J, DBA/2J, C3H/HeJ; haplotype2: NZO/HILtJ,LP/J; haplotype3: FVB/NJ, NOD/ShiLtJ), and four wild-derived strains all carry divergent sequences [16]. Signal-regulatory protein beta 1 members (SIRPB1) are cell surface glycoproteins expressed in leukocytes, which positively regulate neutrophil transepithelial migration [38]. A CNV in SIRPB1 has been reported in humans to be associated with autoimmune thyroid diseases [39].
and impulsive-disinhibited personality [40]. In the GRCm38 reference genome, the Sirpb1 locus remains incomplete [16]. However, the de novo assembly of other inbred mouse strains, especially C57BL/6N, has partially improved the reference genome and confirmed significant conservation and high diversity across the strains compared to the C57BL/6J haplotype at the Sirpb1 locus [16]. Thus, the newly published draft genomes of multiple mouse strains will further facilitate the use of the house mouse for studying human disease.

Both mice and humans carry very large interferon inducible GTPases (GVIN). Their open reading frame is almost 8,000 base pair in length, encoded by a single colossal exon. They are highly expressed in lymph nodes and whole blood in humans [41] and are inducible by both type I and type II interferons (IFNs) in mice [42]. Although the function of GVIN members remains unknown, they are thought to play a role in pathogen immunity [42]. Alpha-1 antitrypsin (AAT) encoded by gene serine protease inhibitor A1 (SERPINA1) is the most abundant antiprotease in humans. It inhibits neutrophil elastase and regulates serine proteases during acute inflammatory responses, especially in the lungs where it protects the fragile alveolar tissues from proteolytic degradation [43]. Human SERPIN family members are highly polymorphic with 1/2,500 newborns in Western Europe carrying the PiZ or PiS allele that causes acute or chronic lung and liver disease [44]. The trade-off of these adverse alleles is still unclear; however, pathogens are reported to manipulate immunity regulation of host as evasion strategies, and SERPINA is a potential target [45]. Among mouse strains, both the SerpinA and SerpinB gene families are highly polymorphic, and mouse strains with different Serpin haplotype may confer a good model for human AAT diversity.

Many more immune-related genes or gene families are encoded in the SSDRs in mice, e.g., IFNs, guanylate-binding proteins (GBP), Ly6 members, orosomucoids (Orm), paired-Ig-like receptor (Pira/Pirb), interferon-induced proteins with tetratricopeptide repeats (IFIT), and CD200 receptors. A summary of these loci can be found in S1 Table.

**Sensory and kin selection**

Key to rodent survival is the ability to detect and avoid potentially harmful compounds by smell and taste. The polymorphisms of Tas2r members are believed to match the profiles of bitter chemicals that the mouse population encounter in their diets. Signatures of positive selection have been detected for the human bitter-taste receptor TAS2R16 [46]. The majority of bitter-taste receptors encoded on mouse chromosome 6 are lineage-specific [47]. Indeed, variations in aversion to chemical substances were observed in BXD mice [48] and between mouse strains C3HeB/FeJ and SWR/J [49]. Both phenotypes were mapped to mouse Tas2r loci on chromosome 6 [50].

Olfactory receptors (ORs) are the largest gene superfamily in house mice and most vertebrates [51]. There are 1,296 OR genes distributed in 27 clusters on the Celera mouse genome except chromosome 12 and Y [51]. As one of the most ancient animal senses, olfaction is important to recognise food, identify mates and offspring, and avoid predators or chemical dangers. Polymorphism in ORs in inbred mouse strains is well studied. Multiple OR members from strains 129S1/SvJ, 129X1/SvJ, 129S6/SvEvTac, A/J, AKR/J, BALB/c, C57BL/6, and DBA/2J were amplified from genomic DNA [51,52] and sequences available [53]. The de novo genome assemblies, especially of wild-derived inbred strains, have greatly boosted the abundance of novel OR genes. Taking strain CAST/EiJ as an example, 1,249 OR candidates have been annotated, 37 of which are not present in the reference mouse genome. In addition, multiple OR pseudogenes in GRCm38 are conserved with CAST/EiJ and vice versa [16]. Strain-specific polymorphisms can be found in 23 OR clusters among 16 mouse strains sequenced. Similar polymorphisms can be found in Taar7d, Taar7e, Taar8a, Taar8b and Taar8c in wild-
derived mouse strains. These members are reported as ORs to recognise ethological odors [54]. The human genome encodes 950 OR genes with high diversity, comparable to the mouse genome [55].

The mouse genome contains many other lineage-specific gene family expansions compared to humans. Many of these genes are associated with reproduction, possibly caused by mating competition and kin selection [56]. One remarkable example is of vomeronasal receptors (VRs) that are mainly expressed in the vomeronasal organ and believed to detect pheromones for sexual recognition. Based on structural differences, VRs are classified into two superfamilies, Vmn1r and Vmn2r, and sum to more than 360 members encoded as clusters on multiple chromosomes in GRCm38 reference genome [57]. The dynamic evolution of VRs and the driving force behind it have been largely discussed in the last decades [58–62]. Wynn and colleagues [57] interrogated around 50% VR genes/alleles from 17 inbred mouse strains and found a significantly higher coding sequence variation with nonrandom distribution in the VRs, especially among three house mouse subspecies and between \( M. \text{Musculus} \) and \( M. \text{Spretus} \). These results suggest that VRs may contribute to reproductive isolation between closely related subspecies [57].

As ligands for VRs [63], the major urinary proteins (Mups) are a set of 18–19 kDa communication proteins abundant in mouse urine and other secretions, including lacrimal, parotid, submaxillary, sublingual, preputial, and mammary glands [64,65]. Mups may either directly behave as pheromones or bind small molecule pheromones to stabilize them by a slow-release pattern [66,67]. In the house mouse, Mups are encoded by a gene-dense cluster on chromosome 4 with at least 19 Mup members per haplotype. Wild mice are reported to express complex ‘barcode’ patterns of Mups, which may provide gender, social dominance, and kinship information to other individuals, facilitating inbreeding avoidance and aiding pup identification. However, wild type individual variation on Mup locus is thought to have been lost during derivation of the classical laboratory strains [68]. It was proposed Mup alleles are also highly conserved between individual wild mice [69]. However, previous research based on PCR amplification could not assay novel haplotypes with highly diverse Mup members. De novo assemblies of 16 mouse strains have confirmed the sequence diversity in all four wild-derived strains [16].

Another important group of pheromone proteins are exocrine gland secreting peptides (ESPs). They may regulate mouse social behaviours via VR activation. Esp1 is reported to mediate Bruce effect in mice [70], and Esp22 secreted by juvenile mice may inhibit adult male mating behaviour[71]. ESPs are encoded by a gene cluster close to the class I MHC. In the GRCm38 reference genome, 38 Esp members are annotated, in which 14 appear to be pseudogenes [72,73]. Although most members in Esp family have high sequence diversity between mouse strains, their polymorphisms have not been widely reported. In the human genome, Mups and ESPs are not present, and all except five V1R genes are disrupted by deleterious mutations [74].

**Behaviour and neuron development**

Sensory receptors may affect the behaviour of the house mouse directly or indirectly, similar to VRs and ORs [75,76]. The modification of behaviour may also be achieved by regulating the development and connection of neuron cells. For example, protocadherin gamma (Pcdhg) members are encoded in a mouse SSDR on chromosome 18. Many Pcdhg members show high polymorphism among mouse strains. Pcdhga genes are found exclusively in vertebrates and predominantly expressed in the nervous system [77]. They may provide a synaptic address code for neuronal connectivity or a single cell barcode for self-recognition and self-avoidance, and their isoform diversity is necessary for postnatal development of neurons [77].
In humans, male and female specific brain function dimorphisms causing mental impairment have been linked to the X chromosome \[78\]. This is partially related to X-linked lymphocyte-regulated (Xlr) members \[79\]. Xlr3b and 4b are paternally imprinted in the cortex and other brain regions, which regulate the expression of other genes\[79\]. Xlr genes are encoded on rapidly evolving gene clusters. Among 16 mouse strains, very few SSDRs can be found on the X chromosome, but the two Xlr loci give strong signatures of CNVs and novel loci in the wild-derived strains.

One of the most remarkable SSDRs is on chromosome 12 (17–25 mega base pairs [mbp]). This 7 Mbp region encodes hippocalcin-like 1 (Hpcal1) and their homologues that belong to the neuronal calcium sensors. Humans have only single copy of HPCAL1, which is mainly expressed in retinal photoreceptors, neurons, and neuroendocrine cells \[80\]. Knockdown of HPCAL1 in neuroblastoma cells led to impaired neurite outgrowth and inhibited sympathetic neuronal differentiation \[81\]. In house mice, this gene has been duplicated into 50–100 copies. The mouse Hpcal1 complex is extremely repetitive containing several recent duplications of hundreds of kilobases. Current draft de novo assemblies do not accurately represent the Hpcal1 locus in any mouse strain, although it appears that in 12 classical strains at least eight different haplotypes can be observed, and four wild-derived strains contain a further four \[16\]. To date, the function of Hpcal1 homologs and the purpose of the rapid expansion of the loci remains unknown. Further candidate genes that potentially have functions in neuron development and regulation include Mas and related G Protein-Coupled Receptors (Mrgpra), Angiopoietins (Ang), and Neuronal apoptosis inhibitory proteins (Naip) \[82–84\].

### Sexual reproduction and other biology processes

Sexual reproduction is a complex process from gamete recognition to maternal-fetal interaction. Many genes related to sperm-egg interaction show positive selection and polymorphism, which may reflect the evolutionary pressure from species recognition or inbreeding avoidance \[85\]. The a disintegrin and metalloprotease (Adam) gene family are important sperm surface proteins. Rapid evolution can be found within their sperm-egg adhesion domains \[86\]. Three Adam members are found in a SSDR, namely Adam20, Adam25, and Adam26a. The divergence can be only observed in M. Spretus, which indicates a potential role in hybridization avoidance \[16\]. Other sperm specific gene families, however, are polymorphic among classical laboratory mice. Sperm-associated glutamate (E)-rich protein (Speer) members are encoded in a gene-dense cluster on chromosome 5. At least three of them are expressed solely in the adult mouse testis. Speer homologs are not present in most other mammal species including humans \[87\], and their function remains unclear. Female specific genes can be also found in the SSDRs. Pregnancy-specific glycoproteins (Psg) are members of immunoglobulin superfamily. In humans, PSGs may be the most abundant trophoblastic proteins in maternal blood during pregnancy \[88\]. Human PSGs play an essential role in the regulation of maternal immunity, by protecting a fetus from immune responses in case of infection, inflammation, and trauma \[89\]. The polymorphism of Psg members in mice is possibly caused by a combination of immune tolerance and host–pathogen coevolution. Four haplotypes of the Psg complex can be found in classical inbred strains, and all wild-derived mouse strains have novel haplotypes.

Many other candidates in the list of mouse strain-specific diversity genes have various function or unknown function (see S1 Table). Variations in keratins and keratin associated proteins (Krtap) may affect the hair content characteristics of mouse individuals \[90\]. Polymorphisms of Hydroxysteroid sulfotransferase enzymes (Sult) may reflect challenge from chemical Metabolism. Variation of zinc finger proteins (Zfp) are thought to repress transposable elements in an evolutionary arms race \[91\].
Conclusions

Isogenic inbred mice have held a unique position as the key mammalian model in evolutionary, genetics, genomics, and biomedical research for over a century. Sequencing and functional studies have documented the extent of genetic polymorphism residing amongst the strains, both shared and unique to each strain. Genetic variation between mouse strains is not evenly distributed across the genome. In most regions, mice are >99.5% identical, but in SSDRs (around 0.5%–2.8% of the mouse genome), the difference is often higher than interspecies diversity between mouse and rat [13]. This scale of diversity cannot be easily represented using the reference genome with SNPs, indels, and SVs. SSDRs are overrepresented with genes associated with immunity, sensory, sexual reproduction, and behavioral phenotypes [16]. The selective pressures driving diversity and CNV includes host–pathogen coevolution (e.g., red queen hypothesis) [92], kin selection [93], mating preference [94], and even selective sweeps due to strong positive selection [95,96]. Many of these genes have direct orthologues in the human genome and are therefore important for understanding health and disease, drug development, and vaccine development. Multiple well-annotated reference genomes will allow researchers to use the appropriate strain for biological rather than historical reasons.

While this review has focused primarily on the limitations of our knowledge of diversity in protein coding regions of the mouse genome, there are other functional elements in which our knowledge is even more limited, e.g., long noncoding RNAs (ncRNAs); piRNAs; and transcription controlling elements such as promoters, enhancers, silencers, and insulators. Multiple reference quality chromosome sequences will provide the foundation for future mapping studies to interrogate these elements. The dramatic drop in second-generation sequencing costs has resulted in genome-wide catalogs of genetic variants for hundreds of mouse strains, but the process of producing a reference quality genome sequence that includes fully resolved novel haplotypes remains costly. Recent advances in third generation sequencing platforms, such as Pacific Biosciences and Oxford Nanopore, can produce mammalian genomes that are an order of magnitude more contiguous[97]. We expect that the representation of many SSDRs in mouse strains will be greatly improved by third generation sequencing platforms.

Human genome-wide association studies (GWAS) have discovered many loci associated with complex disease and traits. Knowledge from model organisms, combined with fine mapping techniques and functional studies, are used to identify causative genes and mechanisms. Mouse SSDR regions are enriched for genes with disease functions with known orthologs in the human genome. The completion of the mouse pan-genome that incorporates all known genetic variants and novel haplotypes will enable the functional characterization of many unresolved quantitative trait loci (QTLs) associated with human disease.

One interesting question is what the origins of these highly diverse haplotypes in the mouse genome are. To date, only a few of these loci have been studied in detail in both inbred and wild mice. Trachtulec and colleagues [98] constructed a haplotype map of the Hst1 region and H2 haplotypes for five mouse subspecies and found that trans-species SNPs were rare, concluding that the haplotypes are unlikely to have arisen by recombination during inbreeding. Lilue and colleagues[13] studied the polymorphic alleles of IRG proteins in inbred laboratory mice that have also been found in European wild mice, suggesting that these alleles arose prior to inbreeding, whilst other more ancient alleles are shared across mouse subspecies. The combination of multiple reference quality genomes for the primary mouse subspecies and availability of larger numbers of sequenced wild mice from ancestral populations will enable a comprehensive analysis of the origins of all SSDRs.
Supporting information

S1 Table. The gene families, publication identifiers, human orthologs, and mouse gene names for the SSDR regions in the mouse genome. SSDR, strain-specific diversity region. (XLSX)

S1 Data. Coordinates on GRCm38 for the SSDR regions per strain (BED format). SSDR, strain-specific diversity region. (GZ)

References

1. Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT, Festing MF, et al. Genealogies of mouse inbred strains. Nat Genet. 2000; 24: 23–25. https://doi.org/10.1038/71641 PMID: 10615122
2. Taft RA, Davison M, Wiles MV. Know thy mouse. Trends Genet TIG. 2006; 22: 649–653. https://doi.org/10.1016/j.tig.2006.09.010 PMID: 17007958
3. Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002; 420: 520–562. https://doi.org/10.1038/nature01262 PMID: 12466850
4. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011; 477: 289–294. https://doi.org/10.1038/nature10413 PMID: 21921910
5. Yalcin B, Wong K, Agam A, Goodson M, Keane TM, Gan X, et al. Sequence-based characterization of structural variation in the mouse genome. Nature. 2011; 477: 326–329. https://doi.org/10.1038/nature10432 PMID: 21921916
6. van der Weyden L, Adams DJ, Bradley A. Tools for targeted manipulation of the mouse genome. Physiol Genomics. 2002; 11: 133–164. https://doi.org/10.1152/physiolgenomics.00074.2002 PMID: 12464689
7. Yang H, Bell TA, Churchill GA, Pardo-Manuel de Villena F. On the subspecific origin of the laboratory mouse. Nat Genet. 2007; 39: 1100–1107. https://doi.org/10.1038/ng2087 PMID: 17660819
8. Yang H, Wang JR, Didion JP, Buus RJ, Bell TA, Welsh CE, et al. Subspecific origin and haplotype diversity in the laboratory mouse. Nat Genet. 2011; 43: 648–655. https://doi.org/10.1038/ng.847 PMID: 21623574
9. Fischer Lindahl K. On naming H2 haplotypes: functional significance of MHC class lb alleles. Immunogenetics. 1997; 46: 53–62. https://doi.org/10.1007/s002510050242 PMID: 9148789
10. Flaherty L, Elliott E, Tine JA, Walsh AC, Waters JB. Immunogenetics of the Q and TL regions of the mouse. Crit Rev Immunol. 1990; 10: 131–175. PMID: 2076187
11. Guénet JL, Bonhomme F. Wild mice: an ever-increasing contribution to a popular mammalian model. Trends Genet TIG. 2003; 19: 24–31. https://doi.org/10.1016/s0168-9525(02)00007-0 PMID: 12493245
12. Hassan MA, Olijnik A-A, Frickel E-M, Saeij JP. Clonal and atypical Toxoplasma strain differences in virulence vary with mouse sub-species. Int J Parasitol. 2019; 49: 63–70. https://doi.org/10.1016/j.ijpara.2018.08.007 PMID: 30471286
13. Lilue J, Müller UB, Steinfeld T, Howard JC. Reciprocal virulence and resistance polymorphism in the relationship between Toxoplasma gondii and the house mouse. eLife. 2013; 2: e01298. https://doi.org/10.7554/eLife.01298 PMID: 24175088
14. Song Y, Endepols S, Klemann N, Richter D, Matuschka F-R, Shih C-H, et al. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. Curr Biol CB. 2011; 21: 1296–1301. https://doi.org/10.1016/j.cub.2011.06.043 PMID: 21724238
15. Bagot S, Campino S, Penha-Gonçalves C, Pied S, Cazenave P-A, Holmberg D. Identification of two cerebral malaria resistance loci using an inbred wild-derived mouse strain. Proc Natl Acad Sci U S A. 2002; 99: 9919–9923. https://doi.org/10.1073/pnas.152215199 PMID: 12114535
16. Lilue J, Doran AG, Fiddes IT, Abrudan M, Armstrong J, Bennett R, et al. Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. Nat Genet. 2018; 50: 1574–1583. https://doi.org/10.1038/s41588-018-0223-8 PMID: 30275530
17. Doran AG, Wong K, Flint J, Adams DJ, Hunter KW, Keane TM. Deep genome sequencing and variation analysis of 13 inbred mouse strains defines candidate phenotypic alleles, private variation and homozygous truncating mutations. Genome Biol. 2016; 17: 167. https://doi.org/10.1186/s13059-016-1024-y PMID: 27480531
18. Li J, Jiang T, Mao J-H, Balmain A, Peterson L, Harris C, et al. Genomic segmental polymorphisms in inbred mouse strains. Nat Genet. 2004; 36: 952–954. https://doi.org/10.1038/ng1417 PMID: 15322544

19. Cutler G, Marshall LA, Chin N, Baribault H, Kassner PD. Significant gene content variation characterizes the genomes of inbred mouse strains. Genome Res. 2007; 17: 1743–1754. https://doi.org/10.1101/gr.6754607 PMID: 17989247

20. Locke MEO, Mijovic M, Eitutis ST, Patel N, Wishart AE, Daley M, et al. Genomic copy number variation in Mus musculus. BMC Genomics. 2015; 16: 497. https://doi.org/10.1186/s12864-015-1713-z PMID: 26140161

21. Morgan AP, Holt JM, McMullan RC, Bell TA, Clayshulte AM-F, Didion JP, et al. The evolutionary fates of a large segmental duplication in mouse [Internet]. Genetics; 2016 Mar. https://doi.org/10.1101/gr.6754607 PMID: 18595730

22. Carlyle JR, Mesi A, Fine JH, Chen P, Bélanger S, Tai L-H, et al. Evolution of the Ly49 and Nkrp1 recognition systems. Semin Immunol. 2008; 20: 321–330. https://doi.org/10.1016/j.smim.2008.05.004 PMID: 18640056

23. Brown MG, Scalzo AA. NK gene complex dynamics and selection for NK cell receptors. Semin Immunol. 2008; 20: 361–368. https://doi.org/10.1016/j.smim.2008.06.004 PMID: 18640056

24. Nobuhara H, Kuida K, Furutani M, Shiroishi T, Moriwaki K, Yanagi Y, et al. Polymorphism of T-cell receptor genes among laboratory and wild mice: diverse origins of laboratory mice. Immunogenetics. 1989; 30: 405–413. https://doi.org/10.1007/bf02421171 PMID: 2574156

25. Barstad P, Farnsworth V, Weigert M, Cohn M, Hood L. Mouse immunoglobulin heavy chains are coded by multiple germ line variable region genes. Proc Natl Acad Sci U S A. 1974; 71: 4096–4100. https://doi.org/10.1073/pnas.71.10.4096 PMID: 4215076

26. Green R, Wilkins C, Thomas S, Sekine A, Hendrick DM, Voss K, et al. Oas1b-dependent Immune Transcriptional Profiles of West Nile Virus Infection in the Collaborative Cross. G3 Bethesda Md. 2017; 7: 1665–1682. https://doi.org/10.1534/g3.117.016124 PMID: 28592649

27. Nakaya Y, Lilue J, Stavrou S, Moran EA, Ross SR. AIM2-Like Receptors Positively and Negatively Regulate the Interferon Response Induced by Cytosolic DNA. mBio. 2017; 8. https://doi.org/10.1128/mBio.00944-17 PMID: 28679751

28. Mavrommatis E, Fish EN, Platanias LC. The schlafen family of proteins and their regulation by interferons. J Interferon Cytokine Res Off J Int Soc Interferon Cytokine Res. 2013; 33: 206–210. https://doi.org/10.1089/jir.2012.0133 PMID: 23570387

29. Sastalla I, Crown D, Masters SL, McKenzie A, Leppla SH, Moayeri M. Transcriptional analysis of the three Nlrp1 paralogs in mice. BMC Genomics. 2013; 14: 188. https://doi.org/10.1186/1471-2164-14-188 PMID: 23506131

30. Shanahan MT, Tanabe H, Ouellette AJ. Strain-specific polymorphisms in Paneth cell a-defensins of C57BL/6 mice and evidence of vestigial myeloid a-defensin pseudogenes. Infect Immun. 2011; 79: 459–473. https://doi.org/10.1128/IAI.00996-10 PMID: 21041494

31. Pemberton AD, Knight PA, Gamble J, Colledge WH, Lee J-K, Pierce M, et al. Inner BALB/c enteric epithelial responses to Trichinella spiralis: inducible expression of a novel goblet cell lectin, intelectin-2, and its natural deletion in C57BL/10 mice. J Immunol Baltim Md 1950. 2004; 173: 1894–1901. https://doi.org/10.4049/jimmunol.173.3.1894 PMID: 15265922

32. Lu ZH, di Domenico A, Wright SH, Knight PA, Whitelaw CBA, Pemberton AD. Strain-specific copy number variation in the intelectin locus on the 129 mouse chromosome 1. BMC Genomics. 2011; 12: 110. https://doi.org/10.1186/1471-2164-12-110 PMID: 21324158

33. Vanhamme L, Paturiaux-Hanocq F, Poelvoorde P, Nolan DP, Lins L, Van Den Abbeele J, et al. Apolipoprotein L-I is the trypanosome lytic factor of human serum. Nature. 2003; 422: 83–87. https://doi.org/10.1038/nature01461 PMID: 12621437

34. Barbee SD, Woodward MJ, Turchinovich G, Mention J-J, Lewis JM, Boyden LM, et al. Skint-1 is a highly specific, unique selecting component for epidermal T cells. Proc Natl Acad Sci U S A. 2011; 108: 3330–3335. https://doi.org/10.1073/pnas.1010890108 PMID: 21300860

35. Boyden LM, Lewis JM, Barbee SD, Bas A, Girardi M, Hayday AC, et al. Skint1, the prototype of a newly identified immunoglobulin superfamily gene cluster, positively selects epidermal gammadelta T cells. Nat Genet. 2008; 40: 656–662. https://doi.org/10.1038/ng.108 PMID: 18408721

36. Percopo CM, Dyer KD, Ochuruk SI, Luo JL, Fischer ER, Lee JJ, et al. Activated mouse eosinophils protect against lethal respiratory virus infection. Blood. 2014; 123: 743–752. https://doi.org/10.1182/blood-2013-05-502443 PMID: 24297871

37. Nitto T, Dyer KD, Mejia RA, Byström J, Wynn TA, Rosenberg HF. Characterization of the divergent eosinophil ribonuclease, mEar 6, and its expression in response to Schistosoma mansoni infection in vivo. Genes Immun. 2004; 5: 668–674. https://doi.org/10.1038/sj.gene.6364143 PMID: 15526002
38. Liu Y, Soto I, Tong Q, Chin A, Bühring H-J, Wu T, et al. SIRPbeta 1 is expressed as a disulfide-linked homodimer in leukocytes and positively regulates neutrophil transepithelial migration. J Biol Chem. 2005; 280: 36132–36140. https://doi.org/10.1074/jbc.M506419200 PMID: 16081415

39. Jin X, Guan Y, Shen H, Pang Y, Liu L, Jia Q, et al. Copy Number Variation of Immune-Related Genes and Their Association with Iodine in Adults with Autoimmune Thyroid Diseases. Int J Endocrinol. 2018; 2018: 1705478. https://doi.org/10.1155/2018/1705478 PMID: 29713342

40. Kajimoto N, Kirpekar SM, Wakade AR. An investigation of spontaneous potentials recorded from the smooth-muscle cells of the guinea-pig seminal vesicle. J Physiol. 1972; 224: 105–119. https://doi.org/10.1113/jphysiol.1972.sp009883 PMID: 5039969

41. Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biol. 2006; 7 Suppl 1: S12.1–14. https://doi.org/10.1186/gb-2006-7-s1-s12 PMID: 16925834

42. Klamp T, Boehm U, Schenk D, Pfeffer K, Howard JC. A giant GTPase, very large inducible GTPase-1, is inducible by IFNs. J Immunol Baltim Md 1950. 2003; 171: 1255–1265. https://doi.org/10.4049/jimmunol.171.3.1255 PMID: 12874213

43. Bergin DA, Hurley K, McElvaney NG, Reeves EP. Alpha-1 antitrypsin: a potent anti-inflammatory and potential novel therapeutic agent. Arch Immunol Ther Exp (Warsz). 2012; 60: 81–97. https://doi.org/10.1016/j.micinf.2017.01.003 PMID: 28093320

44. Lossow K, Hübner S, Roudnitsky N, Slack JP, Pollastro F, Behrens M, et al. Comprehensive Analysis of Mouse Bitter Taste Receptors Reveals Different Molecular Receptive Ranges for Orthologous Receptors in Mice and Humans. J Biol Chem. 2016; 291: 15358–15377. https://doi.org/10.1074/jbc.M116.718544 PMID: 27226572

45. Bouthier JD, Raghow S, Nelson TM, Munger SD. Inbred mouse strains C57BL/6J and DBA/2J vary in sensitivity to a subset of bitter stimuli. BMC Genet. 2005; 6: 36. https://doi.org/10.1186/1471-2156-6-36 PMID: 15967025

46. Nelson TM, Munger SD, Bouthier JD. Taste sensitivities to PROP and PTC vary independently in mice. Chem Senses. 2003; 28: 695–704. https://doi.org/10.1093/chemse/bjg062 PMID: 14627538

47. Zhang X, Firestein S. The olfactory receptor gene superfamily of the mouse. Nat Neurosci. 2002; 5: 124–133. https://doi.org/10.1038/nn800 PMID: 11802173

48. Young JM, Trask BJ. Different evolutionary processes shaped the mouse and human olfactory receptor gene families. Hum Mol Genet. 2002; 11: 535–546. https://doi.org/10.1093/hmg/11.5.535 PMID: 11875048

49. Cassis KJ, Logan DW. Genomic variation in the vomeronasal receptor gene repertoires of inbred mice. BMC Genomics. 2012; 13: 415. https://doi.org/10.1186/1471-2164-13-415 PMID: 22908939
58. 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. https://doi.org/10.3389/fnana.2014.00153 PMID: 25565978

59. Emes RD, Beatson SA, Ponting CP, Goodstadt L. Evolution and comparative genomics of odorant- and pheromone-associated genes in rodents. Genome Res. 2004; 14: 591–602. https://doi.org/10.1101/gr.1940604 PMID: 15060000

60. Lane RP, Young J, Newman T, Trask BJ. Species specificity in rodent pheromone receptor repertoires. Genome Res. 2004; 14: 603–608. https://doi.org/10.1101/gr.2117004 PMID: 15060001

61. Grus WE, Zhang J. Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats. Gene. 2004; 340: 303–312. https://doi.org/10.1016/j.gene.2004.07.037 PMID: 15475172

62. Park SH, Podlaha O, Grus WE, Zhang J. The microevolution of V1r vomeronasal receptor genes in mice. Genome Biol Evol. 2011; 3: 401–412. https://doi.org/10.1093/gbe/evr039 PMID: 21551350

63. Krieger J, Schmitt A, Löbel D, Gudermann T, Schultz G, Brer H, et al. Selective activation of G protein subtypes in the vomeronasal organ upon stimulation with urine-derived compounds. J Biol Chem. 1999; 274: 4655–4662. https://doi.org/10.1074/jbc.274.8.4655 PMID: 9988702

64. Gubits RM, Lynch KR, Kulkarni AB, Dolan KP, Gresik EW, Hollande P, et al. Differential regulation of alpha 2u globulin gene expression in liver, lacrimal gland, and salivary gland. J Biol Chem. 1984; 259: 12803–12809. PMID: 6208189

65. Shahan K, Denaro M, Gilmartin M, Shi Y, Derman E. Expression of six mouse major urinary protein genes in the mammary, parotid, sublingual, submaxillary, and lacrimal glands and in the liver. Mol Cell Biol. 1987; 7: 1947–1954. PMID: 3600653

66. Hurst null, Robertson null, Tolladay null, Beynon null. Proteins in urine scent marks of male house mice extend the longevity of olfactory signals. Anim Behav. 1998; 55: 1289–1297. https://doi.org/10.1006/anbe.1997.0650 PMID: 9632512

67. Chamero P, Marton TF, Logan DW, Flanagan K, Cruz JR, Saghatelian A, et al. Identification of protein pheromones that promote aggressive behaviour. Nature. 2007; 450: 899–902. https://doi.org/10.1038/nature05997 PMID: 18064011

68. Cheetham SA, Smith AL, Armstrong SD, Beynon RJ, Hurst JL. Limited variation in the major urinary proteins of laboratory mice. Physiol Behav. 2009; 96: 253–261. https://doi.org/10.1016/j.physbeh.2008.10.005 PMID: 18973768

69. Thoß M, Enk V, Yu H, Miller I, Luzynski KC, Balint B, et al. Diversity of major urinary proteins (MUPs) in wild house mice. Sci Rep. 2016; 6: 38378. https://doi.org/10.1038/srep38378 PMID: 27922085

70. Hattori T, Osakada T, Masaoka T, Ooyama R, Horio N, Mogi K, et al. Exocrine Gland-Secreting Peptide 1 Is a Key Chemosensory Signal Responsible for the Bruce Effect in Mice. Curr Biol CB. 2017; 27: 3197–3201.e3. https://doi.org/10.1016/j.cub.2017.09.013 PMID: 29033330

71. Ferrero DM, Moeller LM, Osakada T, Horio N, Li Q, Roy DS, et al. A juvenile mouse pheromone inhibits sexual behaviour through the vomeronasal system. Nature. 2013; 502: 368–371. https://doi.org/10.1038/nature12579 PMID: 24089208

72. Kimoto H, Sato K, Nodari F, Haga S, Holy TE, Touhara K. Sex- and strain-specific expression and vomeronasal activity of mouse ESP family peptides. Curr Biol CB. 2007; 17: 1879–1884. https://doi.org/10.1016/j.cub.2007.09.042 PMID: 17935991

73. Kimoto H, Haga S, Sato K, Touhara K. Sex-specific peptides from exocrine glands stimulate mouse vomeronasal sensory neurons. Nature. 2005; 437: 898–901. https://doi.org/10.1038/nature03433 PMID: 16208374

74. Young JM, Massa HF, Hsu L, Trask BJ. Extreme variability among mammalian V1r gene families. Genome Res. 2010; 20: 10–18. https://doi.org/10.1101/gr.08913.109 PMID: 19952141

75. Ibarra-Soria X, Levitin MO, Logan DW. The genomic basis of vomeronasal-mediated behaviour. Mamm Genome Off J Int Mamm Genome Soc. 2014; 25: 75–86. https://doi.org/10.1007/s00335-013-9463-1 PMID: 23884334

76. Glinka ME, Samuels BA, Diodato A, Teillon J, Feng Mei D, Shykind BM, et al. Olfactory deficits cause anxiety-like behaviors in mice. J Neurosci Off J Soc Neurosci. 2012; 32: 6718–6725. https://doi.org/10.1523/JNEUROSCI.4287-11.2012 PMID: 22573694

77. Chen WV, Alvarez FJ, Lefebvre JL, Friedman B, Nwakeze C, Geiman E, et al. Functional significance of isoform diversification in the protocadherin gamma gene cluster. Neuron. 2012; 75: 402–409. https://doi.org/10.1016/j.neuron.2012.06.039 PMID: 22884324

78. Zechner U, Wilda M, Kehrer-Sawatzki H, Vogel W, Fundele R, Hameister H. A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? Trends Genet TIG. 2001; 17: 697–701. https://doi.org/10.1016/s0168-9525(01)02446-5 PMID: 11718922
79. Davies W, Isles A, Smith R, Karunadasa D, Burrmann D, Humby T, et al. Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. Nat Genet. 2005; 37: 625–629. https://doi.org/10.1038/ng1577 PMID: 15908950
80. Burgoyne RD. Neuronal calcium sensor proteins: generating diversity in neuronal Ca2+ signalling. Nat Rev Neurosci. 2007; 8: 182–193. https://doi.org/10.1038/nrn2093 PMID: 17311005
81. Wang W, Zhong Q, Teng L, Bhatnagar N, Sharma B, Zhang X, et al. Mutations that disrupt PHOXB interaction with the neuronal calcium sensor HPCAL1 impede cellular differentiation in neuroblastoma. Oncogene. 2014; 33: 3316–3324. https://doi.org/10.1038/onc.2013.290 PMID: 23873030
82. Geppetti P, Veldhuis NA, Lieu T, Bunnell NW. G Protein-Coupled Receptors: Dynamic Machines for Signaling Pain and Itch. Neuron. 2015; 88: 635–649. https://doi.org/10.1016/j.neuron.2015.11.001 PMID: 26590341
83. Subramanian V, Crabtree B, Acharya KR. Human angiogenin is a neuroprotective factor and amyotrophic lateral sclerosis associated angiogenin variants affect neurite extension/pathfinding and survival of motor neurons. Hum Mol Genet. 2008; 17: 130–149. https://doi.org/10.1093/hmg/ddm290 PMID: 17916583
84. Götz R, Karch C, Digby MR, Troppmaier J, Rapp UR, Sendtner M. The neuronal apoptosis inhibitory protein suppresses neuronal differentiation and apoptosis in PC12 cells. Hum Mol Genet. 2000; 9: 2479–2489. https://doi.org/10.1093/hmg/9.17.2479 PMID: 11030753
85. Firman RC, Gasparini C, Manier MK, Pizzari T. Postmating Female Control: 20 Years of Cryptic Female Choice. Trends Ecol Evol. 2017; 32: 368–382. https://doi.org/10.1016/j.tree.2017.02.010 PMID: 28318651
86. Cvjetta A. Positive selection within sperm-egg adhesion domains of fertilin: an ADAM gene with a potential role in fertilization. Mol Biol Evol. 2003; 20: 21–29. https://doi.org/10.1093/molbev/msg002 PMID: 12519902
87. Spiess A-N, Walther N, Müller N, Balvers M, Hansis C, Ivell R. SPEER—a new family of testis-specific genes from the mouse. Biol Reprod. 2003; 68: 2044–2054. https://doi.org/10.1095/biolreprod.102.011593 PMID: 12606357
88. Moore T, Dveksler GS. Pregnancy-specific glycoproteins: complex gene families regulating maternal-fetal interactions. Int J Dev Biol. 2014; 58: 273–280. https://doi.org/10.1387/ijdb.130329gd PMID: 25023693
89. Motrán CC, Díaz FL, Gruppi A, Slavin D, Chatton B, Bocco JL. Human pregnancy-specific glycoprotein 1a (PSG1a) induces alternative activation in human and mouse monocytes and suppresses the accessory cell-dependent T cell proliferation. J Leukoc Biol. 2002; 72: 512–521. PMID: 12223519
90. Wu D-D, Irwin DM, Zhang Y-P. Molecular evolution of the keratin associated protein gene family in mammals, role in the evolution of mammalian hair. BMC Evol Biol. 2008; 8: 241. https://doi.org/10.1186/1471-2148-8-241 PMID: 18721477
91. Imbeault M, Helleboi d P-Y, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature. 2017; 543: 543–544. https://doi.org/10.1038/nature21683 PMID: 28273063
92. Morrano, Schmidt OG, Gelarden IA, Parrish RC, Lively CM. Running with the Red Queen: Host-Parasite Coevolution Selects for Biparental Sex. Science. 2011; 333: 216–218. https://doi.org/10.1126/science.1206360 PMID: 21737739
93. Axelrod R, Hammond RA, Grafen A. Altruism via kin-selection strategies that rely on arbitrary tags with which they coevolve. Evol Int J Org Evol. 2004; 58: 1833–1838.
94. Sherborne AL, Thom MD, Paterson S, Jury F, Ollier WER, Stockley P, et al. The genetic basis of inbreeding avoidance in house mice. Curr Biol CB. 2007; 17: 2061–2066. https://doi.org/10.1016/j.cub.2007.10.041 PMID: 17993707
95. Didion JP, Morgan AP, Clayshulte AM-F, McMullan RC, Yadgary L, Petkov PM, et al. A multi-megabase copy number gain causes maternal transmission ratio distortion on mouse chromosome 2. PLoS Genet. 2015; 11: e1004850. https://doi.org/10.1371/journal.pgen.1004850 PMID: 25679959
96. Didion JP, Morgan AP, Yadgary L, Bell TA, McMullan RC, Ortiz de Solorzano L, et al. R2d2 Drives Selfish Sweeps in the House Mouse. Mol Biol Evol. 2016; 33: 1381–1395. https://doi.org/10.1093/molbev/msw036 PMID: 26882987
97. Gordon D, Huddleston J, Chaisson MJ, Hill CM, Kronenberg ZN, Munson KM, et al. Long-read sequence assembly of the gorilla genome. Science. 2016; 352: aae0344. https://doi.org/10.1126/science.aee0344 PMID: 27034376
98. Trachtulec Z, Vlcek C, Mihola O, Gregorova S, Fotopulosova V, Forejt J. Fine Haplotyp e Structure of a Chromosome 17 Region in the Laboratory and Wild Mouse. Genetics. 2008; 178: 1777–1784. https://doi.org/10.1534/genetics.107.082404 PMID: 18245833