Received: 4 April 2019 | Revised: 10 June 2019 | Accepted: 12 June 2019

DOI: 10.1111/mec.15155

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Evolutionary patterns of major urinary protein scent signals in house mice and relatives

Michael J. Sheehan[1] | Polly Campbell[2] | Caitlin H. Miller[1]

[1]Neurobiology and Behavior, Cornell University, Ithaca, NY, USA

[2]Evolution, Ecology and Organismal Biology, University of California – Riverside, Riverside, CA, USA

Correspondence
Michael J. Sheehan, Neurobiology and Behavior, Cornell University, Ithaca, NY, USA.
Email: msheehan@cornell.edu

Abstract

Scent marks are important mediators of territorial behaviour and sexual selection, especially among mammals. The evolution of compounds used in scent marks has the potential to inform our understanding of signal evolution in relation to social and sexual selection. A major challenge in studies of chemical communication is that the link between semiochemical compounds and genetic changes is often unclear. The major urinary proteins (MUPs) of house mice provide information on sex, status and individual identity. Importantly, MUPs are a direct protein product of genes, providing a clear link between genotype and phenotype. Here, we examine the evolution of urinary protein signals among house mice and relatives by examining the sequences and patterns of mRNA expression of *Mup* genes related to urinary scent marks. MUP patterns have evolved among mouse species both by gene duplication and variation in expression. Notably, protein scent signals that are male specific in well-studied inbred laboratory strains vary in sex-specificity among species. Our data reveal that individual identity signals in MUPs evolved prior to 0.35 million years ago and have rapidly diversified through recombining a modest number of amino acid variants. Amino acid variants are much more common on the exterior of the protein where they could interact with vomeronasal receptors, suggesting that chemosensory perception may have played a major role in shaping MUP diversity. These data highlight diverse processes and pressures shaping scent signals, and suggest new avenues for using wild mice to probe the evolution of signals and signal processing.

KEYWORDS

commensalism, domestication, pheromones, rodent, social evolution, social neuroscience

1 | INTRODUCTION

Scent marks provide complex blends of socially salient information that are important drivers of behavioural interactions for many animal species (Ferkin, 2019; Gosling & Roberts, 2001; Hurst & Beynon, 2004). Depending on the species and source of the scent, markings may contain information about species identity, sex, age, health, physiological state, kinship, group membership, and individual identity (Wyatt, 2010, 2014). Information in scent marks may reflect the genetic makeup (Charpentier, Crawford, Boulet, & Drea, 2010; Mateo, 2003; Sheehan et al., 2016) as well as the behavioural, sexual and social state of an individual (Ferris, Axelson, Shinto, & Albers, 1987; Freeman, Sheehan, & Ophir, 2019; Martín & Lopez, 2007). Given the importance of scent communication for many species, there has been considerable interest in the evolutionary patterns of scent mark composition and information content across species. Pheromone diversification within and between species has been of particular interest (Fang et al., 2009; Janssenswillen et al., 2014; Lassance et al., 2010; Tupec et al., 2019) because pheromones mediate social and sexual interactions within species, and potentially

act as a prezygotic barriers between populations (Ganem, Rueff, & Perriat-Sanguinet, 2014).

House mice are highly territorial scent markers. As the principal mammalian model organism (Phifer-Rixey & Nachman, 2015), house mice are an important model for understanding the structure and function of scent marks (Desjardins, Maruniak, & Bronson, 1973; Hurst, 1987; Thonhauser, Raveh, Hettyey, Beissmann, & Penn, 2013). Both laboratory and wild house mice use urine to mark territories. These territorial urine marks are subsequently used to assess competitors and potential mates. Work on genetically diverse wild populations of house mice (*Mus musculus domesticus*) has consistently demonstrated a critical role for major urinary proteins (MUPs) in scent marking (Cheetham et al., 2007; Green et al., 2015; Hurst & Beynon, 2013). MUPs are involatile lipocalin proteins that act directly as pheromones, and also influence volatile components of urine markings (Chamero et al., 2007; Nevison, Armstrong, Beynon, Humphries, & Hurst, 2003; Timm, Baker, Mueller, Zidek, & Novotny, 2001). The mouse genome encodes more than 20 tandemly arrayed *Mup* genes, a subset of which are expressed highly in the liver and excreted in the urine at high concentrations (Logan, Marton, & Stowers, 2008; Mudge et al., 2008). House mice produce two different classes of urinary MUPs with distinct expression profiles. The first class of urinary MUPs are known as central MUPs, due to their physical position within the tandem gene array. Males and females express a diversity of highly similar central MUP isoforms. While the total amount of central MUPs produced varies with social conditions (Nelson, Cunningham, Ruff, & Potts, 2015), the specific pattern of MUPs is genetically determined and fixed for a given individual (Sheehan et al., 2016). Importantly, the number and identity of central MUPs varies among individuals within a population, allowing MUP signatures to serve as a signal of individual identity (Cheetham et al., 2007; Hurst et al., 2001; Sheehan et al., 2016). Individual recognition of scent marks via central MUPs is used in male-male competition, female assessment of males, and for kin recognition (Green et al., 2015; Hurst, 2009; Ramm, Cheetham, & Hurst, 2008). In addition to their role in individual recognition, variation in central MUPs allows mice to avoid mating with close relatives, cooperate with kin, and assess heterozygosity (Green et al., 2015; Thom et al., 2008). The second class of urinary MUPs or peripheral MUPs, which include *Mup3* and *Mup20*, have been characterized as male-specific pheromones in common laboratory strains (Mudge et al., 2008; Roberts et al., 2010). Due to their sex-biased expression, these MUPs appear to indicate whether a scent mark has been made by a male or female. *Mup3* has been shown to stimulate male-male countermarking and aggression (Kaur et al., 2014). *Mup20* (also known as darcin) stimulates learning about spatial locations of male scent marks in females and promotes countermarking in males (Kaur et al., 2014; Roberts, Davidson, McLean, Beynon, & Hurst, 2012). Additionally, it has been shown to stimulate aggression to intruder males by lactating females (Martín-Sánchez et al., 2015).

The MUP signatures in wild mouse populations (*M. m. domesticus*) provide important social information on individual identity and sex. Previous analyses of mammalian genomes and rodent urine have shown that the patterns of MUP production seen in house mice are not widespread among rodents (Beynon et al., 2008; Hagemeyer et al., 2011; Logan et al., 2008). This raises interesting questions about the evolutionary origins of information content in mouse MUP signatures. Specifically, we address four questions:

1. *What information is encoded by MUP blends?* Current evidence from *M. m. domesticus* suggests that MUP blends encode information about individual identity, sex, and male dominance status. Information content of signals is determined by the distribution of variation among individuals within and between groups (Dale, 2006; Tibbetts, Mullen, & Dale, 2017). Individual identity information is encoded by the diverse central MUPs, genes that differ among individuals in both sequence and relative expression patterns (Hurst et al., 2001, 2017; Sheehan et al., 2016). Sex is encoded by the differential expression of *Mup3* and *Mup20*. Males are reported to express both genes while females do not (Mudge et al., 2008). Furthermore, higher status males increase expression of *Mup20* relative to other MUPs in their urine (Thoß et al., 2019). If semiochemicals are specific to a given species or population, they may also encode information about the species identity or population of origin (Mullen, Mendelson, Schal, & Shaw, 2007). Here, we examine patterns of expression in males and females to assess the potential for MUPs to encode individual identity and sex across species. MUP signatures that encode individual identity could show either of two patterns, both contributing to individually recognizable signatures. First, individuals may differ in the amino acid sequences of proteins that are excreted in their urine. Second, individuals can also vary in the relative levels of the same proteins. Similarly, MUP signatures encoding sex should show a bias in relative expression patterns between sexes of the same species.
2. *Do specific proteins retain the same information across species?* In order for protein signatures to encode relevant social or sexual information there needs to be variation in expression among individuals or sexes. However, which proteins vary need not be the same. For example, male biased expression of a protein in one species versus female biased expression of the same protein in another, both provide information on sex. However, the information encoded by the presence of the protein differs between species. Indeed, work in *Drosophila* has shown that sex differences in expression of key pheromones differs across species, such that the information content of a particular molecule varies across the phylogeny (Seeholzer, Seppo, Stern, & Ruta, 2018).
3. *Are pheromones shared or species-specific?* In addition to encoding sex and individual information, MUPs in scent marks may encode species identity through two mechanisms (Symonds & Elgar, 2008). First, the same compounds may be used in distinct ratios or blends to indicate species. Alternatively, amino acid differences between proteins may allow differential detection and response to otherwise similar protein ratios. Proteins that are shared across species may indicate functional constraints on either MUP

structure or the structure of receptors. Different proteins across species may indicate selection for divergent protein pheromone repertoires across species. Because many semiochemicals are the products of multiple genes (Lassance et al., 2010, 2013), linking divergence in phenotype to function changes in genotype is a significant challenge. In the case of the mice, however, MUPs are direct gene products allowing for clear assessment of predicted protein similarity from transcriptomic sequences.

4. *What are the long-term consequences of selection for individual identity signatures?* Theory and experiments suggests that traits under selection for individuality should be under negative frequency-dependent as rare phenotypes are more recognizable and therefore facilitate correct identification (Dale, Lank, & Reeve, 2001; Johnstone, 1997; Sheehan, Miller, & Reeve, 2017; Sheehan & Tibbetts, 2009). While much of individuality in scent marks has been assumed to be the consequence of broadscale genetic and physiological differences among individuals (Todrank & Heth, 2003; Willse et al., 2006), results from mice indicate that MUPs specifically encode individual identity (Hurst et al., 2001; Kaur et al., 2014; Roberts et al., 2018). Indeed, molecular analyses of *Mup* genes within a population of *M. m. domesticus* identified clear signatures of negative frequency-dependent selection (Sheehan et al., 2016). The longer-term consequences of selection for individual identity signatures on patterns of phenotypic diversification have yet to be explored. Two possible outcomes from negative frequency-dependent selection have been observed. On the one hand, selection may maintain particular genetic and phenotypic variants over long periods of time. For example, negative frequency-dependent selection maintains multiple male mating strategies for long periods of time in lineages such as *Uta* lizards (Corl, Davis, Kuchta, & Sinervo, 2010; Sinervo, Bleay, & Adamopoulou, 2001). On the other hand, selection for individuality may not depend on a few common variants maintained over long periods of time, but rather on the generation of novel variants. High diversity and turnover of alleles used in some fungal mating systems are indicative of such a process (James, 2012). We can begin to tease apart these possibilities by examining the extent to which different species, subspecies, or populations share particular protein variants.

Here, we combine newly generated and publicly available transcriptomic data to examine the sequences and patterns of *Mup* genes expressed in the liver of house mice and close relatives. While MUP scent signaling ultimately depends on protein excretion in the urine, examination of mRNA sequences from gene expression data allow for broad insights into the evolution of MUP scent marks because of the wide availability of liver transcriptomes. Previous work has already established that mice and rats have independently expanded their *Mup* gene families (Gomez-Baena et al., 2018; Logan et al., 2008). Therefore, we focus our efforts on species in the genus *Mus* for which liver transcriptomic sequences are available.

2 | MATERIALS AND METHODS

2.1 | Generation of liver transcriptomes

We harvested liver tissues from adult males and females for laboratory mice from multiple strains and species representing approximately 7 million years of evolution within the genus *Mus*. Tissues were harvested from adult mice kept in breeding cages within a standard mouse colony. RNA was extracted from tissue using a Qiagen RNeasy Kit. RNA sequencing libraries were generated using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB #E7530). NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490) was used for RNA Isolation. Sequences were indexed using the NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1, NEB #E7600). We sequenced paired-end libraries at the Institute for Biotechnology at Cornell University. Some libraries were sequenced on a MiSeq (PE 300) while others were sequenced on a NextSeq (PE 150). Information for specimens used in for sequencing is given in Table S1: Appendix S1. All procedures were approved by the Institutional Animal Care and Use Committee of Cornell University. Newly sequenced data is available under BioProject: PRJNA530260.

2.2 | Criteria for choosing publicly available liver transcriptomes for analysis

We searched the NCBI Short Read Archive (https://www.ncbi.nlm.nih.gov/sra) for liver transcriptomes of muroid rodents with the goal of identifying transcriptional data for species in *Mus* and related genera. The vast majority of sequences are derived from experiments examining differential expression in liver transcriptomes across treatments in laboratory mouse strains, which we did not consider. Rather, we focused our attention on data from unmanipulated liver samples that provided information on the expression of urinary proteins across a diverse group of mice. In total, we examined 100 samples from 10 species or subspecies (Figure 1, Table S1: Appendix S1).

2.3 | De novo liver transcriptomes

For species that are more distantly related to the house mouse and thus more likely to have poor sequence alignment with the reference genome (*M. caroli*, *M. pahari*, *M. mattheyi and Apodemus uralensis*) we generated de novo liver transcriptomes to identify potential *Mup* sequences. RNAseq reads generated for this study (*M. caroli* and *M. pahari*) or downloaded from the short read archive (*M. mattheyi and A. uralensis*) were assembled using default settings in rnaSPAdes (Bushmanova, Antipov, Lapidus, & Przhibelskiy, 2018). We then queried known *Mup* transcript sequences from the house mouse against assembled transcriptomes to identify *Mup* transcripts using BLAST. All identified transcripts appeared to be full length *Mup* genes upon manual alignment and inspection.

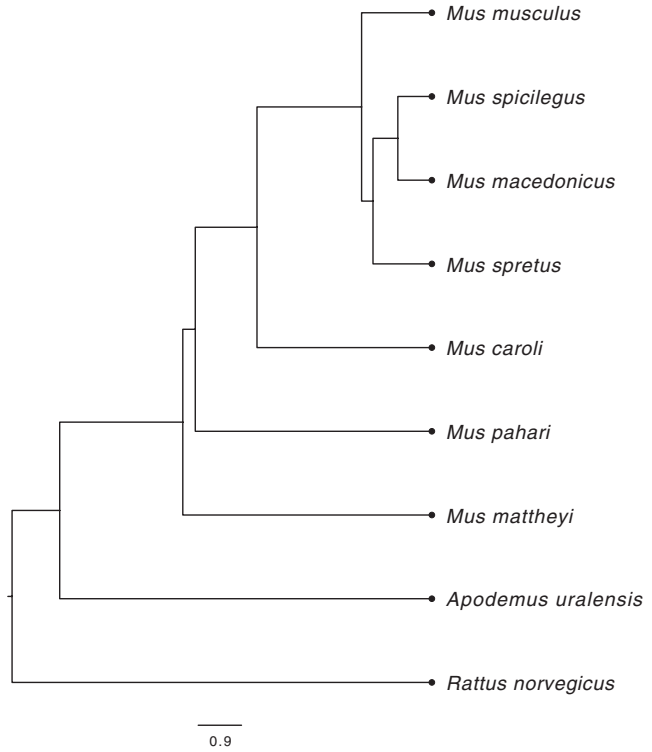

FIGURE 1 Species tree for muroid rodents examined in this study. Modified from (Steppan & Schenk, 2017). The scale bar measures substitutions per site

2.4 | Alignment of RNASeq reads

In *M. m. domesticus*, alignment of short reads from MUP genes or transcripts is problematic due to the high similarity of sequences within the central MUP cluster. Previous work examining MUP diversity (Sheehan et al., 2016) aligned sequences to a modified transcriptome consisting of sequences of *Mup3*, *Mup11* and *Mup20*. Both *Mup3* and *Mup20* are examples of peripheral *Mup* genes that located on the edge of the gene cluster and have distinct sequences. In contrast, *Mup11* is a central *Mup* gene that is one of many similar genes found on in the centre of the gene cluster. For distantly related species *M. caroli*, *M. pahari* and *M. mattheyi* we also added sequences derived from de novo transcriptomes. We did not detect any *Mup* genes in the transcriptome of *A. uralensis* so no sequences were added for this species. Reads were aligned with BOWTIE2 using the "–very-fast" setting.

2.5 | Comparing patterns of gene expression

Urinary protein output varies among individuals both in the overall production of MUPs and the relative production of each protein within an individual (Nelson et al., 2015; Sheehan et al., 2016). In comparing gene expression, our goal was to understand the relative contributions to pheromone blends of different MUP types across individuals and species. Visual inspection of read alignments to MUP transcripts showed that coverage varied along the transcripts in relation to the sequencing strategy employed. The shorter reads used in some of the Short Read Archive experiments produced more even coverage along the transcript than PE300 sequencing experiments, though longer sequences allowed for easier identification of full-length transcripts. To account for differences in sequencing strategy and depth across all the samples we scored the level of expression among genes as the deepest read count for each gene. From these measures we then calculated the relative contributions of each gene type (central *Mup* sequences aligned to *Mup11*, sequences aligned to *Mup3*, and sequences aligned *Mup20*) to the predicted composite protein blend for each individual. This approach has the benefit of capturing within-individual allocation of MUPs, rather than variation in MUP expression relative to other proteins in the liver as a whole, as would be the case in a typical differential expression analysis. Previous work utilizing the same approach for documenting relative *Mup* transcript abundance showed that this method predicts the final patterns of relative protein excretion in a wild population of *M. m. domesticus* at a nearly 1:1 ratio (Sheehan et al., 2016). Moreover, the tight correlation between protein excretion and liver transcripts suggests that these measures probably provide an accurate view of the relative abundance of different MUPs excreted in the urine. We note that while we have confidence that our methods are accurately capturing the variation in *Mup* gene expression in the samples examined here it is important to keep in mind that the samples we have used for analysis in many cases have been taken from the Short Read Archive. Thus, our study has not been specifically designed to control for potential effects of social conditions, age, etc on levels of *Mup* gene expression. Nevertheless, despite the noise introduced by the use of heterogenous samples from the Short read archive, we detect clear patterns of expression across strains and species.

2.6 | Searching for *Mup* genes in genome assemblies

We searched for *Mup* genes in published genomes for the *M. m. musculus* strain PWK/PhJ, the *M. m. castaneus* strain CAST/EiJ, and *M. spretus*, *M. spicilegus*, *M. caroli* and *M. pahari* (Couger, Arévalo, & Campbell, 2018; Keane et al., 2011; Thybert et al., 2018). All genomes except *M. spicilegus* were accessed via the Ensemble genome browser 95. The genome for *M. spicilegus* was downloaded from GenBank. Transcript sequences for the *Mup11*, *Mup3* and *Mup20* were downloaded from the Mouse Genome Informatics website (http://www.informatics.jax.org/). For all species, we queried the transcripts of the three genes against the genome using BLAST. We used the conserved exon/intron structure among *Mup* genes within the mouse genome to identify full length gene sequences in each assembly. Within the BLAST results we searched for consecutive strings of sequence covering the whole transcript over approximately 3–5 kb. The entire sequences encompassing the transcript ±1 kb were then downloaded and aligned with other known *Mup* genes using the online version of MAFFT v.7 (https://mafft.cbrc.jp/alignment/software/) and then further examined by hand using

Mesquite (Maddison, 2008). Only sequences predicted to produce a full length protein were retained. To classify sequences, we generated a gene tree of all predicted mRNA sequences using IQ-Tree (Nguyen, Schmidt, von Haeseler, & Minh, 2014). We generated a ML tree with 1,000 bootstraps. The best evolutionary model was selected using the built-in model selection function. In this case the best mode was: TMP3+F+G4. Gene sequences identified are provided in Appendix S3.

2.7 | Bespoke detection of variants and assembly of full length predicted mature MUP sequences

The high sequence similarity among central MUPs makes the automated assembly of high-confidence *Mup* gene sequences challenging. With paired end RNAseq data, however, it is possible to reconstruct the sequences of transcript sequences associated with mature MUPs (162 amino acids). Using the Integrated Genome Viewer v 2.4 (Robinson et al., 2011), we determined full-length sequences of predicted mature MUP using the following strategy. First, we adjusted the allele cut-off for showing variants to 0.05 for central MUPs to account for the fact that in some individuals 10–20 central MUPs may be excreted. For pooled samples, this process probably misses some MUP sequences but allows for high confidence in those that are generated. For the peripheral MUPs, we kept the cut-off value at 0.3. Reads were viewed as pairs and sorted by insert size (largest inserts first). For each variant within the sequence, we sorted by base so that all variants containing a particular nucleotide were sorted together. Combinations of variants were recorded. This process was repeated for each variable site. Duplicate sequences generated by this process were noted and a final output of unique sequences for each sample was generated. Gene sequences used for analysis are provided in Appendix S2.

Due to relatively lower *Mup* gene expression levels in females, we limited this hand-done analysis to male samples, where higher average levels of gene expression facilitated confident gene assemblies. As a result, we make no attempt to analyze sex-specific patterns of expression among individual central *Mup* genes, as opposed to the aggregate level of central *Mup* gene expression, for which we can have high confidence. There is evidence for sex-biased expression of the central *Mup* gene, *Mup7*, in the inbred laboratory strain C57BL6/J (Mudge et al., 2008). Thus, it is important to note the limitation of this study with regard to analysis fo sex-specificity of any given central *Mup* gene.

2.8 | Molecular diversity of mature central MUPs

mRNA sequences for each MUP were translated to proteins using standard RNA-DNA conversions in Mesquite. We then aligned all central MUP genes from this study as well as those previously identified from wild *M. m. domesticus* from Edmonton, Canada (Sheehan et al., 2016) to examine the distribution of amino acid variants across the proteins and their relative abundance among the proteins sampled. The 3D structure of MUP11 is a barrel of beta-sheets, characterized by an inner hydrophobic pocket that binds and transports volatile compounds, and a hydrophilic exterior that is assumed to interact with MUP-detecting vomeronasal receptors Phelan, McLean, Hurst, Beynon, and Lian (2014). The model includes the relative position of each amino acid on the interior or exterior of the protein, noted as the percentage exposure to the exterior (broken into quintiles), or as part of the central barrel. Thus, we counted the number of proteins that possess a nonreference Mup11 amino acid at each position relative to how exposed it is to the exterior.

2.9 | Gene tree of *Mup* sequences

A gene tree of *Mup* transcripts corresponding to mature proteins excreted in the urine was generated from sequences aligned using the online version of MAFFT v.7 in IQTree using the same procedure as above. A single *Mup* gene sequence from the independent *Mup* gene family expansion in *Rattus* (ENSRNOT00000046760) was used as an outgroup.

3 | RESULTS

3.1 | Sequences of *Mup* genes expressed in the liver

Though both house mice and rats express a diversity of *Mup* genes in their livers, other mouse species and strains examined here showed wide variation in the number and types of *Mup* genes expressed (Figure 2). While we detected *Mup* genes in all *Mus* species sampled we were unable to detect any *Mup* orthologs in the liver transcriptome of *A. uralensis*, indicating that the extent of liver expression of *Mup* genes is probably quite variable among even close relatives of house mice and rats. Details of all of the *Mup* gene sequences and predicted proteins surveyed in this study are provided in Tables S1 and S2: Appendix S1.

With the exception of the *Mup* gene expressed by *M. mattheyi*, the detected *Mup* genes can be readily classified as being central MUPs or related to *Mup3* or *Mup20* (Figure 2). The gene expressed by *M. mattheyi* appears to be basal to the subsequent diversification of central MUPs and *Mup20* in other species. Subspecies of *M. musculus* tend to express a single conserved version of *Mup3* and *Mup20* but other species have at least one amino acid variant differentiating sequences between species. Notably, the close relatives of house mice either express *Mup3* (*M. spicilegus* and *M. macedonicus*) or *Mup20* (*M. spretus*) but not both.

3.2 | Patterns of *Mup* gene expression in the liver vary widely across mouse species

Across species and sexes, we identified three key ways that *Mup* gene expression varies (Figure 3): (a) the level of overall investment in *Mup* gene expression, (b) which types of *Mup* genes are expressed, and (c) the relative ratios of different types of *Mup* genes. The three

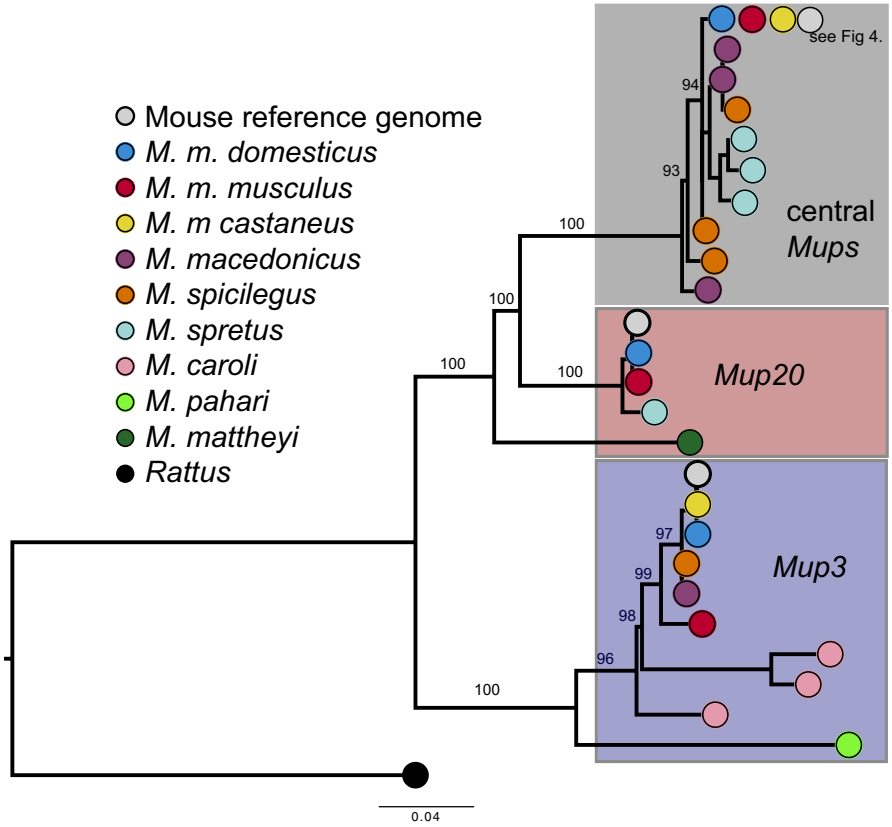

FIGURE 2 Simplified gene tree of *Mup* genes expressed in the liver. Expressed sequences for each species or subspecies are shown with a different colour (see legend). Reference sequences for *Mup11*, *Mup3*, *Mup20* and an outgroup sequence for a rat MUP has been used to root the tree. For the three subspecies of *M. musculus* only data for one inbred laboratory strain each are shown here. The diversity of central MUPs in *M. musculus* laboratory and wild samples are represented here with *Mup11* sequence. See Figure 4 for additional details on central *Mup* gene diversity. The scale bar shows substitutions per site. Bootstrap values are show for major nodes. Full details in Appendix S4

patterns variously contribute to differences in *Mup* gene expression between sexes and among species. Combined with sequence differences in predicted MUPs (Figure 2), the patterns of expression suggest that MUP profiles provide information about species and sex.

Investment in overall levels of *Mup* expression varies among species and sexes (ANOVA, species: $F_{7,89}$ = 10.05, p = 5.25e−9; sex: $F_{1,89}$ = 47.13, p = 1.17e−9, species*sex: $F_{7,89}$ = 2.8, p = .012). Across species, males show higher levels of *Mup* gene expression (Figure 3), consistent with patterns previously reported in subspecies of *M. musculus* (Hurst et al., 2017; Stopková, Stopka, Janotova, & Jedelsky, 2007). Overall levels of expression vary across species, though expression levels in males and females are correlated (linear regression, t_6 = 3.16, r^2 = .63, p = .019). There is considerably higher investment in *Mup* expression in *M. musculus* subspecies males compared to other species, especially *M. spicilegus*, *M. pahari* and *M. mattheyi*, which have low levels of *Mup* gene expression (Figure 3). It is worth noting that even in these species with "low" expression levels *Mup* genes are in fact highly expressed compared to other genes in the liver transcriptome. Levels of expression in female *M. spicilegus* and *M. pahari* are exceedingly low, however, suggesting that very little protein is likely to be found in their urine. A lack of *Mup* gene expression was also seen in male *M. spicilegus* from the wild-derived strain SPI/TUA, though males of other wild-derived *M. spicilegus* strains as well as wild-caught males expressed modest levels of *Mup* transcripts.

Patterns of expression across species and sexes revealed that *Mup3* is not a male specific gene. Female *M. m. domesticus*, *M. m. castaneus* and *M. macedonicus* all express *Mup3* at modest levels (Figure 3). Male *M. m. musculus* express a *Mup3*-like gene at very low levels (predicted to be «1% of total MUP content; Figure 3). This same pattern of expression is present in laboratory strains PWK/PhJ and CZECHII/EiJ as well as pooled samples from wild-caught male *M. m. musculus*, indicating that the low expression of an unusual *Mup3*-like gene is not an artifact of inbreeding in wild-derived strains. Our analysis also reveals that *M. caroli* has a lineage-specific expansion of *Mup3* related genes (Figures 2 and 3). Two of these genes are expressed by both males and females while a third is strongly male-biased.

With the notable exception of *M. m. domesticus*, expression of *Mup20* appears to be limited to species or subspecies that do not express *Mup3* (or only express it at very low levels). Both *M. mattheyi* and *M. spretus* express genes related to *Mup20* but not *Mup3* (Figures 2 and 3). The expression of *Mup20* is very low, however, in *M. spretus*. Similarly, *M. pahari*, *M. caroli*, *M. macedonicus*, *M. spicilegus* and *M. m. castaneus* all express orthologs of *Mup3* but not of *Mup20*. Male *M. m. musculus* express *Mup20* at high levels (Figure 3) but, as mentioned above, express a *Mup3*-like gene at extremely low levels. In both *M. m. musculus* and *M. m. domesticus* expression of *Mup20* is strongly male-biased, consistent with previous reports (Hurst et al., 2017; Roberts et al., 2010).

Expression of central *Mup* genes is absent in *M. mattheyi*, *M. pahari*, and *M. caroli* but is the dominant form of *Mup* expressed in the livers of *M. spretus*, *M. spicilegus*, *M. macedonicus* and *M. musculus* subspecies (Figure 3). This pattern holds for both males and females. In fact, both *M. spretus* and *M. m. musculus* females appear to only express central *Mup* genes.

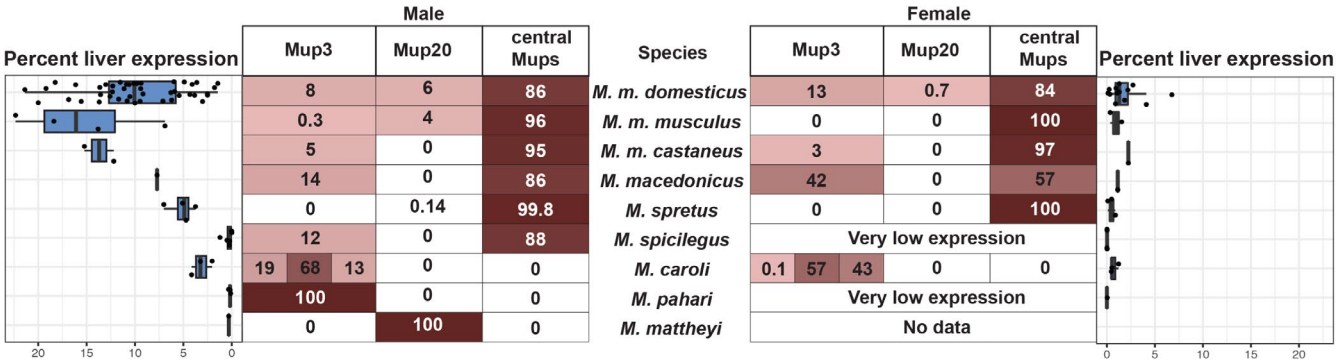

FIGURE 3 Distribution of *Mup* gene expression among species and between sexes. The average percentages of total *Mup* gene expression composed of *Mup3*, *Mup20* and central MUPs are shown for males and females of each species (where data are available). High proportions are shown as darker red. The box plots on the side of the chart show the level of total *Mup* expression, calculated as the percent of all reads that aligned to *Mup* genes, for each sample examined. The three different figures reported for Mup3 in *M. caroli* represent each of the three Mup3-like genes detected in the urine of that species

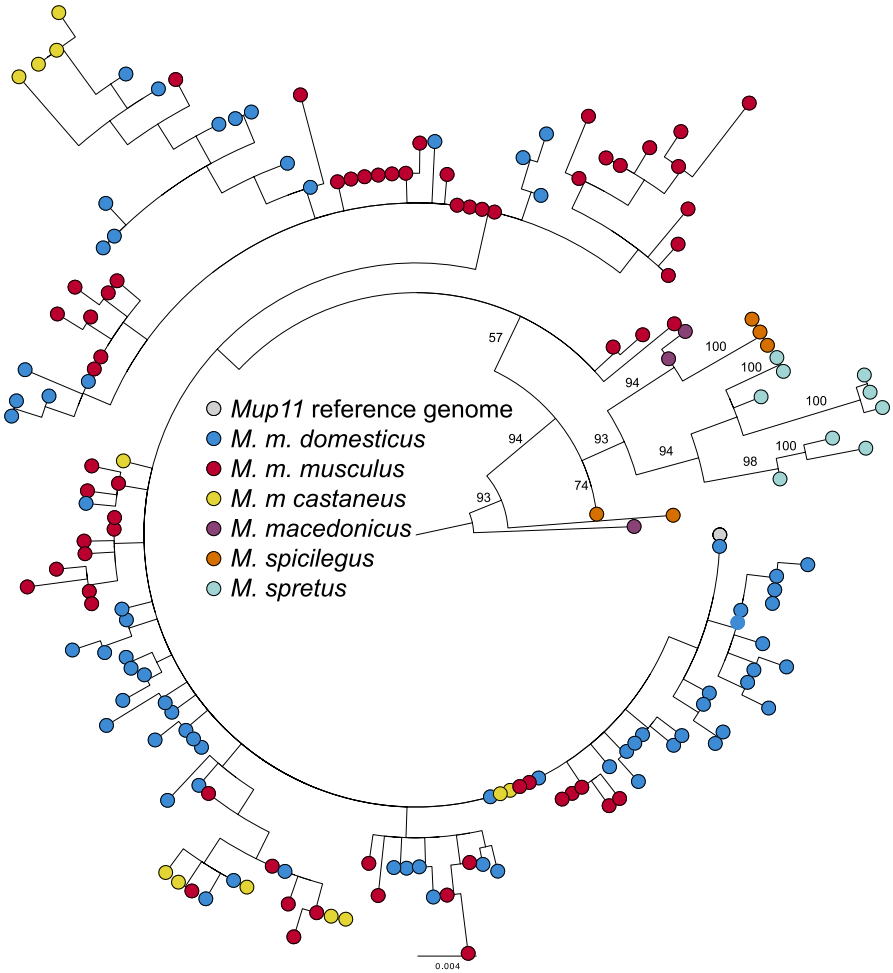

FIGURE 4 Gene tree of central *Mup* genes expressed in the livers of mice. Expressed sequences for each species or subspecies are shown with a different colour (see legend). The gene tree indicates that the large expansion of central *Mup* genes previously reported for *M. m. domesticus* is shared by other *M. musculus* subspecies, suggesting that the expansion predates subspecies divergence approximately 0.35 million years ago. The scale bar shows substitutions per site. Bootstrap values are shown for major nodes. Full details for bootstrap values are provided in Appendix S4

3.3 | Genomic evidence for *Mup* genes not expressed in the liver

The variation in expression of *Mup* gene types (3, 20 and central) across species could arise either because of the differential gain and loss of genes among species, through differential gene regulation, or some combination of the two. To test for the role of gene gains and losses, we examined genomic data for *M. pahari*, *M. caroli*, *M. spretus*, *M. spicilegus*, *M. m. musculus* (PWK/PhJ) and *M. m. castaneus* (CAST/EiJ). In all cases, the *Mup* gene cluster is poorly assembled, as is also the case for the mouse reference genome (Mudge et al., 2008). Therefore, we limited our query to whether orthologs of *Mup3*,

Mup20 and a central MUP could be recovered from genomic data for cases in which a transcript was not detected in the liver. A maximum likelihood gene tree was estimated to determine the orthology relationships for genes identified in the examined genomes (Figure S1, Appendix S3).

We detected genomic evidence for genes that are not transcribed in the liver in five of the genomes examined (Table 1). We did not detect any full-length *Mup* genes in the *M. spicilegus* genome assembly, although this is probably an issue with assembly rather than a true lack of genes since we detect expression of full-length *Mup* transcripts (Figures 2 and 3). Both the *M. m. castaneus* and *M. caroli* genome assemblies contain single copy orthologs for the *Mup* genes missing from liver transcriptomes. Multiple genes in *M. pahari* were identified that are basal to the divergence of *Mup20* and central MUPs clade and may be orthologs to the liver-expressed *Mup* gene identified in *M. matheyi*. However, weak resolution in this part of the gene tree precludes strong conclusions regarding orthology (Figure S1, Appendix S3). The *M. spretus* genome assembly contains evidence for multiple peripheral *Mup* genes though there is no detectable ortholog for *Mup3*. While consistent with *Mup3* loss, this finding could be an artifact of fragmentary assembly of the tandemly-arrayed *Mup* gene family.

3.4 | Multiple expansions of central *Mup* genes

The diversity of central *Mup* genes detected in the liver transcriptomes of *M. spretus*, *M. spicilegus*, *M. macedonicus* and *M. musculus* subspecies is shown in detail in Figure 4. Relative to the other mouse species examined here, all of these species show an expanded repertoire of central *Mup* genes, indicating that there was a gene expansion in the common ancestor of this clade. We detected three different predicted proteins expressed in *M. macedonicus* and an additional three in *M. spicilegus*, though neither forms a species-specific monophyletic clade. Previous work examined the MUPs in urine of *M. macedonicus* and found a single dominant protein with a molecular mass of 18,742 Da (Robertson, Hurst, Searle, Gunduz, & Beynon, 2007). The predicted dominant protein from the liver transcriptome is expected to have a molecular mass of 18,739 Da, which is similar but outside the range of measurement error, suggesting differences between the wild samples and the wild-derived inbred laboratory strain. We did, however, identify a protein with mass 18,742 Da in *M. spicilegus*. In two wild-derived strains, ZRU and ZBN, this single protein is the only central MUP predicted for males and it is the dominant MUP predicted for a wild-caught sample.

A small expansion of sequences found in wild-derived and wild-caught *M. spretus* form a monophyletic clade nested within the *M. spicilegus* and *M. macedonicus* sequences. Previous work examining the variation in MUPs in wild *M. spretus* urine identified three proteins (Beynon et al., 2008). We have identified nine different genes sequences that are predicted to produce six distinct proteins. In the wild-derived strain SPRET/EiJ we found only three proteins. The most abundant protein has a predicted mass of 18,758 Da, similar to the most abundant protein reported from wild-caught samples. The other predicted proteins from SPRET/EiJ have similar masses to previously reported proteins measured from wild *M. spretus* (18,668 and 18,685 Da vs. previously reported 18,666 and 18,687 Da from Beynon et al., 2008).

The large expansion of central *Mup* sequences found among the *M. musculus* subspecies form a monophyletic group nested within the larger central *Mup* clade, although the bootstrap value of 57 indicates poor support at the node leading to *M. musculus* central *Mup* genes. The nodes within the *M. musculus* clade generally have very poor support (Appendix S4). Weak phylogenetic signal may be caused by a combination of purifying selection on the overall gene sequences, and gene conversion (Sheehan et al., 2016).

Due to the poor resolution of the central *Mup* genes in the *M. musculus* clade it is difficult to draw many firm conclusions, though two patterns appear from the data. First, there are many genes found in each subspecies. This is even true for *M. m. castaneus*, represented by just two wild-derived inbred strains (CAST/EiJ and TWN). Second, parts of the expansion are shared in common across the three subspecies. One gene is shared among all three and at least two other clades feature sequences from all three subspecies. Shared expansions across all three subspecies are consistent either with a scenario in which central *Mup* genes underwent an initial expansion prior to subspecies diversification, or there has been introgression of central *Mup* genes among subspecies.

TABLE 1 Genomic evidence for nonexpressed *Mup* genes

Species	*Mup3*	*Mup20*	Central MUPs
M. m. domesticus	T	T	T
M. m. musculus	T	T	T
M. m. castaneus	T	Genomic	T
M. spretus	No evidence	T	T
M. caroli	T	Genomic	Genomic
M. pahari	T	Genomic[a]	

Abbreviation: T, transcriptomic.
[a]These genes are basal to the *Mup20*/central MUPs divergence.

3.5 | Molecular diversity of central MUP proteins

The 165 central *Mup* transcripts detected across *M. musculus* subspecies samples (Figure 4) encode 77 distinct proteins. The 77 protein variants arise from combinations of 39 amino acid variants spread across 32 sites (Figure 5a). We examined the amino acid sequences among *M. musculus* subspecies samples and compared these to amino acid variants found in *M. spretus*, *M. macedonicus* and *M. spicilegus*. Only five of 39 amino acid variants are shared with other species. Thus, most amino acid variants appear to be specific to house mice, the majority of which were detected in only a single sample (50/77 proteins, Table S2, Appendix S5). Please note that this finding does not necessarily indicate that proteins tend to be limited to single individuals, as many of the previously published RNAseq

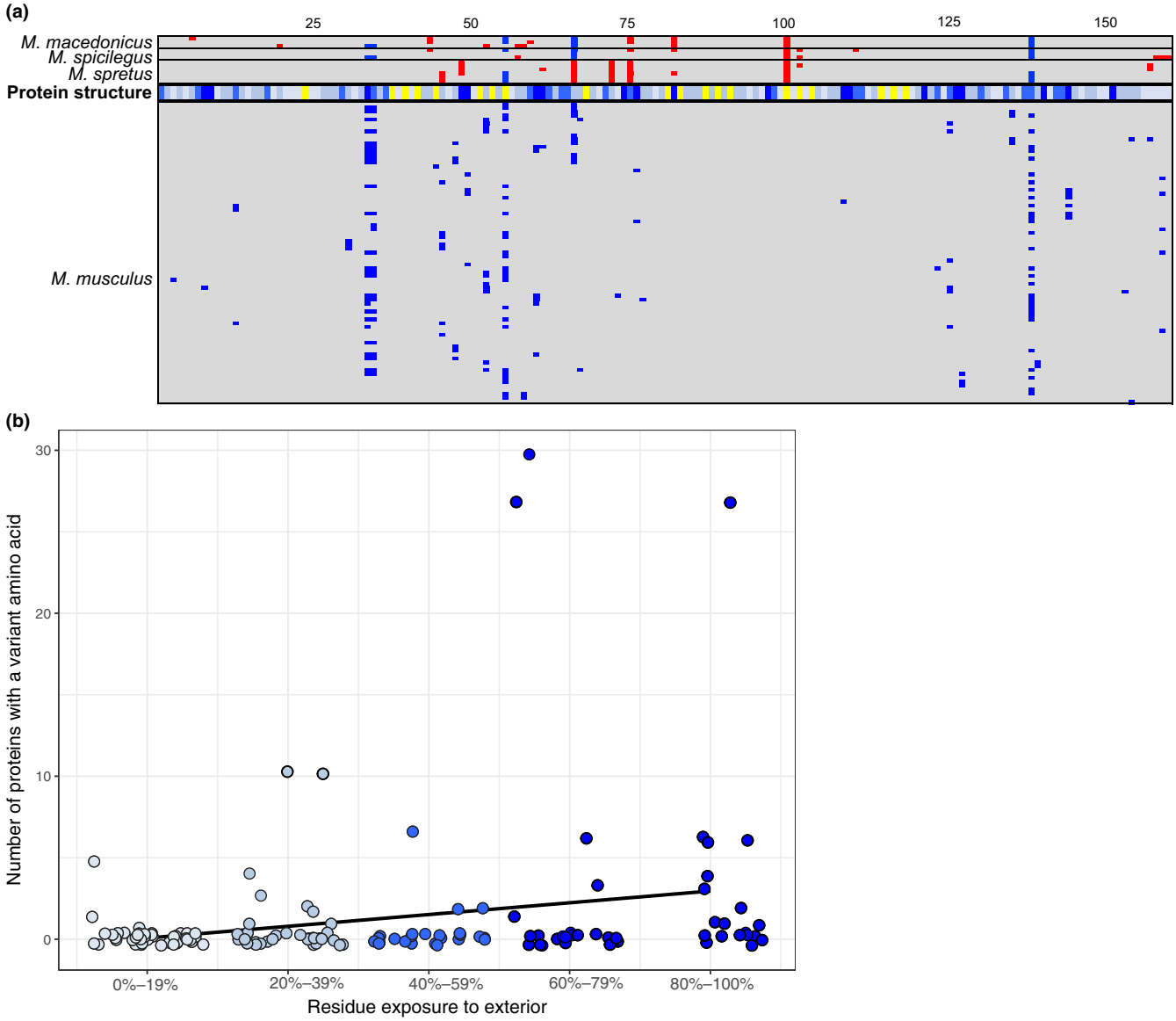

FIGURE 5 Molecular diversity of central MUP proteins. (a) The distribution of amino acid variants among central MUPs is depicted for predicted mature protein sequences (162 amino acids length). Amino acid variants that differ from the sequence of a MUP11 are show in blue if found in a *M. musculus* subspecies, red if only found in a different species. Sequences for *M. spretus*, *M. macedonicus* and *M. spicilegus* are shown at the top while sequences for *M. musculus* subspecies are shown on the bottom. The bar labelled "protein structure" indicates the position of the amino acid resides within the 3D protein structure based on Phelan et al. (2014). Yellow indicates an amino acid is part of the interior beta-sheet barrel. Exposure levels on the exterior of the protein are shown as quintiles (0%–20% lightest blue to 80%–100% darkest blue). Greater detail, including the protein sequences, is provided in Appendix S1. (b) Variants are more common among amino acid sites with greater exposure to the exterior of the protein, which can interact with vomeronasal receptors. The figure shows the number of protein variants containing sites within each category of exposure (excluding sites in the internal barrel). Colouring of points corresponds to the "protein structure" bar in (a), where darkest blue is the highest level of exterior exposure

data set of wild-caught mice re-examined here are pooled samples. Nevertheless, it does indicate that MUP blends composed of both rare and common proteins are probably the norm in many house mouse populations. Of the 27 proteins detected in more than one sample, seven are shared between two subspecies and an additional two are shared by all three subspecies. No proteins were shared across all samples.

We next investigated how protein structure may influence patterns of diversity. Most amino acid variants appear in one or a few proteins, although four variants are found in more than 25 proteins. The variable sites are predominantly found on the hydrophilic exterior surface of the protein (30/32) compared to the interior hydrophobic pocket (2/32). The extent of residue exposure predicts the level of variation observed across sites (Figure 5b, Spearman rank order correlation, $t_{140} = 3.41$, $r_s = .27$, $p = .0008$). Thus, protein variants are more abundant at sites that probably influence binding between MUPs and vomeronasal receptors.

4 | DISCUSSION

Our investigation of the evolutionary dynamics of urinary protein signal production in *Mus* reveals three key findings. First, our results indicate that the information content of MUPs has changed over time. More basal species have single MUPs expressed by males, though the identity of those proteins varies. In more recently derived lineages, both males and females express MUPs though the relative differences in expression patterns vary across species. Second, variation in relative expression between males and females indicates that the information content of a given protein changes across lineages. Specifically, the male specificity of *Mup3* and *Mup20* varies across species, suggesting that the responses to these proteins are also likely to differ across taxa. Third, the extreme diversity in central MUP genes in house mice is specific to the *M. musculus* species group and predates subspecies diversification. Moreover, most proteins in our data set are relatively rare across samples and only a few are shared across subspecies, suggesting that protein diversity evolves and turns over rapidly across house mouse populations. Notably, the result that diversity is overrepresented in sites with greater exterior exposure that may interact with vomeronasal receptors is consistent with the functional diversification of central MUPs as signals of individual identity.

4.1 | Two modes of evolution in mouse major urinary protein scent signals

Phermone blends and identity signatures can evolve through changes in the relative ratios of semiochemicals or via changes to the specific pheromone compounds (Symonds & Elgar, 2008; Wyatt, 2010). Both modes of signal evolution appear to occur in mice. Variation in the identity of MUPs expressed in the liver has been achieved through two different mechanisms: gene family expansion and changes in patterns of gene expression.

There appear to have been multiple gene family expansions followed by sequence divergence contributing to differences in MUP blends across species. At least four gene expansions within *Mus* are well-supported by present data. However, additional expansions seem likely; better assembled sequences of the entire *Mup* gene cluster across species will provide a clearer picture of gene family evolution. Previous examinations of *Mup* gene family evolution have indicated an independent expansion within *Mus* relative to *Rattus* (Logan et al., 2008). The present analysis provides further details on the relative timing and nature of at least four expansion events (Figures 2, 4 and S1).

1. At least one initial expansion event appears to have occurred relatively early within *Mus*, as indicated by the diversity of genes present in the *M. pahari* genome and the fact that *M. mattheyi* and *M. pahari* appear to express different *Mup* genes in their livers.
2. An expansion of *Mup3*-like genes appears to have occurred on the lineage leading to *M. caroli* that is absent in other species examined here
3. An initial expansion of central *Mup* genes occurred in the common ancestor of *M. spretus*, *M. spicilegus*, *M. macedonicus* and *M. musculus*. All species appear to have at least three central *Mup* genes though the genes do not fall into three clades with orthologs for each species. Previous studies examining patterns of diversity among central *Mup* genes in a wild population of *M. m. domesticus* found evidence of gene conversion (Sheehan et al., 2016). Given the similarity in the gene sequences among central *Mup* genes within each species, gene conversion may obscure some aspects of the timing of diversification within and between species.
4. Prior to the diversification of the three subspecies of house mouse approximately 350,000 years ago there seems to have been an additional expansion of central *Mup* genes. This expansion has produced individually distinctive MUP blends used for individual recognition (Hurst et al., 2001; Kaur et al., 2014).

Additional variation in urinary pheromone blends has been achieved through differential expression of genes across species. Previous work has demonstrated that within a population of *M. m. domesticus* there is variation in whether or not mice express certain central *Mup* genes (Sheehan et al., 2016). The present study demonstrates that different species of mice typically possess copies of central MUPs, *Mup3* and *Mup20* in their genome that are not expressed in the liver (Figure 3, Table 1). Why are nonexpressed genes maintained in species' genomes? We only examined genes predicted to produce full-length proteins suggesting they are not pseudogenes. Of course, a lack of expression in the liver does not mean the *Mup* genes are not expressed in other tissues. There is evidence that MUP proteins are excreted in bodily fluids other than urine and may serve a signalling function in other secretions (Černá, Kuntová, Talacko, Stopková, & Stopka, 2017; Stopková et al., 2016). Thus, selective pressures on the maintenance and diversification of *Mup* genes is probably driven by social communication not limited to territory marking. Moreover, *Mup3*, *Mup20* and *Mup11* are all expressed in tissues other than the liver (Finger et al., 2016) and may have other yet to be described functions that may favor the maintenance of these genes in the genome.

4.2 | Variation in sex-specificity of MUP3, a presumed "male-specific" pheromone

Previous studies have shown that expression of MUP3 is male-limited in the laboratory strain C57BL/6J (aka B6), leading to the assumption that this is a male-specific protein (Mudge et al., 2008). Behavioural work in B6 has further shown that MUP3 is sufficient to induce males to countermark and promotes aggressive behaviour in the presence of another male (Kaur et al., 2014), as might be expected for a pheromone that signals male ownership of a scent mark. However, we find robust evidence for female expression of *Mup3* in

M. m. domesticus, *M. m. castaneus*, *M. macedonicus*, and of Mup3-like genes in *M. caroli*. While overall *Mup* gene expression is low in female *M. macedonicus*, *Mup3* genes are expressed at relatively high levels compared to central *Mup* genes. Whether expression leads to biologically relevant excretion of Mup3 proteins in female *M. macedonicus* remains to be determined. Female *M. m. domesticus* and *M. m. castaneus*, however, do express reasonably high quantities of *Mup3*, such that the protein would probably be a relevant part of their urinary scent. Though mRNA expression in the liver appears to predict urinary execretion of proteins among *Mus*, it is possible that females do not excrete a MUP3 protein in spite of modest gene expression levels, though that remains to be tested with wild populations. MUP3 does not show up on standard analyses of urine protein content using mass spectrometry due to glycosylation that increases the mass, although it is detectable via other methods (discussed further as B6 gene18 in Mudge et al., 2008). Gel electrophoresis detects expression in male B6 urine but not female. Thus, the peculiar features of *Mup3* and a dearth of transcriptomic analyses have probably led to a misestimation of its typical expression patterns in wild populations.

Comparative analyses of pheromones in *Drosophila* show a similar pattern of variation in sex specificity of key compounds among species (Seeholzer et al., 2018). Changes in sex-limited expression of key pheromones may be a straightforward way for species to diverge in sexual signaling from close relatives where hybridization may be disfavored, as is the case among mouse subspecies (Smadja, Catalan, & Ganem, 2004; Smadja & Ganem, 2002). While these results do not call into question previous behavioural studies on the effects of MUP3 on male behaviour, they do suggest the possibility that neural mechanisms for processing the same semiochemicals have diverged among closely related species of mice, as has been show in *Drosophila* (Seeholzer et al., 2018). Repeating studies of *Mup3* effects on behaviour in wild-derived strains or wild-caught mice from different subspecies will be important to understanding the behavioural selection pressures influencing the evolution of urinary signals in mice.

4.3 | Liver expression of darcin is a recent phenomenon

Darcin (MUP20) is a male-biased protein that is highly expressed in territorial male *M. m. domesticus* (Roberts et al., 2010) and *M. m. musculus* (Thoß et al., 2019). Our data support these previous studies. We detect very minimal expression in male *M. spretus*, and do not detect darcin expression in *M. m. castaneus*, *M. macedonicus* or *M. spicilegus*. Behavioural and neurobiological studies have demonstrated that darcin promotes female *M. m. domesticus* attraction to male scents and memory for the location where those scents were encountered (Roberts et al., 2012). Furthermore, darcin leads to increased countermarking in B6 males (Kaur et al., 2014). Based on patterns of gene expression, we predict that *M. m. musculus* would respond similarly to darcin, whereas response would be different or absent in other closely related species. Recent changes in expression of darcin among closely related species provides an excellent opportunity to examine the mechanisms of species differences in responses to pheromone signals. This is especially true given the rich set of responses documented to darcin (Hoffman, Pickavance, Thippeswamy, Beynon, & Hurst, 2015; Kaur et al., 2014; Roberts et al., 2012, 2010).

4.4 | Individual identity signatures and patterns of pheromone diversification

Scent signatures are commonly used for individual recognition across diverse animal taxa including ants (D'Ettorre & Heinze, 2005), fish (Thünken, Waltschyk, Bakker, & Kullmann, 2009), lizards (Carazo, Font, & Desfilis, 2008), and many mammals (Johnston, 2003; Thom & Hurst, 2004). Traits used for individual recognition are expected to be under negative frequency-dependent selection (Dale et al., 2001; Tibbetts et al., 2017), as this process will maintain diversity in traits needed for differentiating among individuals. Despite wide interest in recognition abilities, little work has examined the molecular diversity and evolution of traits used for recognition, probably because the underlying molecular basis is typically unknown. Because they are directly derived from gene products, MUPs provide an unusual opportunity to investigate the evolution of individual identity. The present study provides key insights in the origins of individual identity and the processes shaping trait diversification.

Our data refine the estimate for the expansion of central MUPs that mediate individual recognition in house mice. Previous work had suggested that central MUP diversification may date to the split between *M. musculus* and *M. spretus* (Mudge et al., 2008) while other authors had suggested high population density associated with human commensalism may have driven the expansion of central *Mup* genes (Logan et al., 2008). The present study indicates that there were two distinct phases of central MUP expansion: a minor expansion in the common ancestor of *M. musculus* and *M. spretus* and a second larger expansion prior to the diversification of *M. musculus* subspecies, which occurred approximately 350,000 years ago (Geraldes et al., 2008). Thus, the dramatic expansion of central MUPs that mediate recognition is more recent than previously estimated but predates commensalism with humans, because commensalism probably evolved independently in the three subspecies. The ability to advertise individual identity in urinary scent marks is likely to be beneficial in dense populations observed in *M. musculus* both in feral and commensal settings (Hurst, 1987; Pocock, Searle, & White, 2004). One tantalizing possibility is that increased identity information in urinary signals facilitated the evolution of commensalism with humans approximately 10,000 years ago. Identity signals may have facilitated the independent evolution of commensalism in each subspecies. High population densities associated with human commensalism in mice should increase interaction rates among animals, making identity information a potentially vital mediator of territory marks. A correlation between social structure and the extent of identity has also been reported for alarm calls in marmots (Pollard

& Blumstein, 2011), indicating that social interactions can favor increased signal individuality. While it is likely that commensalism has further selected for the elaboration and diversification of central *Mup* genes, the data presented here (Figure 4) indicate that the initial elaboration happened prior to subspecies diversification.

We have detected distinct sets of *Mup* mRNA from transcriptome data among strains and wild samples from *M. m. musculus*. These data stand in contrast to one recent study that claimed a lack of diversity in *Mup* haplotypes in a wild population of *M. m. musculus* (Thoß et al., 2016). That previous study, however, used Sanger sequencing of the amplicon of a single exon pooled across all *Mup* genes (Thoß et al., 2016). Such a method would be inherently unlikely to identify differences in relatively rare sites as would be necessary when many individual *Mup* gene variants tend to be rare among the genes found within an individual (Sheehan et al., 2016, this study). Protein analysis of the same *M. m. musculus* population by the same group, however, report variability in the number of distinct protein bands and proteins identified from male urine (Enk et al., 2016; Thoß, Luzynski, Ante, Miller, & Penn, 2015). Indeed, protein heterogeneity among *M. m. musculus* has been further reported by other groups (Hurst et al., 2017) highlighting that MUP diversity appears to be widespread among *M. musculus* subspecies.

Analyses of the MUP isoforms found across populations and subspecies of house mice provide insights into the dynamics of selection on individual identity signaling phenotypes. Consistent with a model of negative frequency-dependent selection favouring novel proteins, we find evidence for rapid turnover of protein composition across *M. musculus* subspecies and populations. Of 77 distinct proteins, only two proteins were shared across all subspecies, and 50 proteins were specific to a single sample (though not necessarily a single individual as some samples in the NCBI Short Read Archive were pools of wild-caught individuals). Given the relatively small sample of mice studied, further sampling is almost certain to uncover many more central MUP isoforms in wild mouse populations.

Compared to rarity of shared proteins across populations, amino acid substitutions are more commonly shared among proteins. Amino acid variants on the exterior of the protein are much more likely to be found in a greater number of protein variants. The existence of a relatively small number of hypervariable sites has two implications for the evolution of MUP protein diversity. First, the observation is consistent with previous analyses suggesting that gene conversion is probably commonplace among central MUPs (Mudge et al., 2008; Sheehan et al., 2016), as they do not segregate into discrete groups but display a mosaic distribution across proteins (Figure 5a). Second, these few highly variable sites are likely candidates for interaction sites between central MUP ligands and their corresponding vomeronasal receptors (Chamero et al., 2007; Kaur et al., 2014). While it is known that MUPs are detected by V2R vomeronasal receptors, which of the hundreds of V2R's in the mouse genome respond to which MUPs are presently unknown. This precludes a detailed analysis of the co-evolution of receptors and ligand at present. Nevertheless, the sharing of a modest number of amino acid variants across proteins may be the result of convergent evolution shaped by a limited set of possible amino acid variants.

4.5 | The utility of wild house mice for understanding pheromone function and evolution

The rapidly expanding neurobiological toolkit in laboratory mice for assessing neural processing can be readily translated to wild house mice and their relatives. The diversity of predicted pheromone blends documented in this study suggests that wild mice and their relatives will be an unusually valuable system for understanding the evolution of complex pheromone-mediated behaviours. Further research is now needed to understand how mice differ in the use of urinary scent marks as well as how they respond to individual components of scent blends. Coupled with the toolset available for mouse genetics and neurobiology, studies of wild mouse scent marks have the potential to substantially further our understanding the role of pheromone signals in behaviour and evolution.

ACKNOWLEDGEMENTS

Funding for this research was provided by Cornell University. The *M. macedonicus* XBS and *M. spicilegus* ZRU strains were originally developed by the Wild Mouse Genetic Repository (University of Montpellier). We thank Tory Hendry for assistance with generating gene trees.

AUTHOR CONTRIBUTIONS

M.J.S. conceived of the study, analyzed the data, and wrote the initial manuscript. P.C. and C.H.M. generated samples and sequence data. All authors contributed to the editing of the manuscript.

DATA ACCESSIBILITY

Newly sequenced liver transcriptomes are available from the NCBI Short Read Archive under BioProject PRJNA530260. Accessions for publicly available data used are provided in Table S1: Appendix S1. Sequences used for analyses are provided in supplemental files.

ORCID

Michael J. Sheehan https://orcid.org/0000-0002-3949-7873
Polly Campbell https://orcid.org/0000-0001-7660-9814

REFERENCES

Beynon, R. J., Hurst, J. L., Turton, M. J., Robertson, D. H., Armstrong, S. D., Cheetham, S. A., … Humphries, R. E. (2008). Urinary lipocalins in Rodenta: Is there a generic model? In J. L. Hurst, R. J. Benyon, S. C. Roberts, & T. D. Wyatt (Eds.), *Chemical signals in vertebrates 11* (pp. 37–49). New York, NY: Springer.

Bushmanova, E., Antipov, D., Lapidus, A., & Przhibelskiy, A. D. (2018). rnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. *bioRxiv*, 420208.

Carazo, P., Font, E., & Desfilis, E. (2008). Beyond 'nasty neighbours' and 'dear enemies'? Individual recognition by scent marks in a lizard (*Podarcis hispanica*). *Animal Behaviour*, *76*, 1953–1963. https://doi.org/10.1016/j.anbehav.2008.08.018

Černá, M., Kuntová, B., Talacko, P., Stopková, R., & Stopka, P. (2017). Differential regulation of vaginal lipocalins (OBP, MUP) during the estrous cycle of the house mouse. *Scientific Reports*, *7*(1), 11674. https://doi.org/10.1038/s41598-017-12021-2

Chamero, P., Marton, T. F., Logan, D. W., Flanagan, K., Cruz, J. R., Saghatelian, A., … Stowers, L. (2007). Identification of protein pheromones that promote aggressive behaviour. *Nature*, *450*, 899–U23. https://doi.org/10.1038/nature05997

Charpentier, M. J., Crawford, J. C., Boulet, M., & Drea, C. M. (2010). Message 'scent': Lemurs detect the genetic relatedness and quality of conspecifics via olfactory cues. *Animal Behaviour*, *80*(1), 101–108. https://doi.org/10.1016/j.anbehav.2010.04.005

Cheetham, S. A., Thom, M. D., Jury, F., Ollier, W. E. R., Beynon, R. J., & Hurst, J. L. (2007). The genetic basis of individual-recognition signals in the mouse. *Current Biology*, *17*, 1771–1777. https://doi.org/10.1016/j.cub.2007.10.007

Corl, A., Davis, A. R., Kuchta, S. R., & Sinervo, B. (2010). Selective loss of polymorphic mating types is associated with rapid phenotypic evolution during morphic speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(9), 4254–4259. https://doi.org/10.1073/pnas.0909480107

Couger, M. B., Arévalo, L., & Campbell, P. (2018). A high quality genome for *Mus spicilegus*, a close relative of house mice with unique social and ecological adaptations. *G3: Genes, Genomes, Genetics*, *8*(7), 2145–2152.

Dale, J. (2006). Intraspecific variation in coloration. In G. E. Hill, & K. J. McGraw (Eds.), *Bird coloration*, *Volume 2*, *Function and evolution*. Cambridge, MA: Harvard University Press.

Dale, J., Lank, D. B., & Reeve, H. K. (2001). Signaling individual identity versus quality: A model and case studies with ruffs, queleas, and house finches. *The American Naturalist*, *158*(1), 75–86. https://doi.org/10.1086/320861

Desjardins, C., Maruniak, J. A., & Bronson, F. H. (1973). Social rank in house mice: Differentiation revealed by ultraviolet visualization of urinary marking patterns. *Science*, *182*(4115), 939–941.

D'Ettorre, P., & Heinze, J. (2005). Individual recognition in ant queens. *Current Biology*, *15*, 2170–2174. https://doi.org/10.1016/j.cub.2005.10.067

Enk, V. M., Baumann, C., Thoß, M., Luzynski, K. C., Razzazi-Fazeli, E., & Penn, D. J. (2016). Regulation of highly homologous major urinary proteins in house mice quantified with label-free proteomic methods. *Molecular Biosystems*, *12*(10), 3005–3016. https://doi.org/10.1039/C6MB00278A

Fang, S., Ting, C.-T., Lee, C.-R., Chu, K.-H., Wang, C.-C., & Tsaur, S.-C. (2009). Molecular evolution and functional diversification of fatty acid desaturases after recurrent gene duplication in Drosophila. *Molecular Biology and Evolution*, *26*(7), 1447–1456. https://doi.org/10.1093/molbev/msp057

Ferkin, M. H. (2019). Scent marks of rodents can provide information to conspecifics. *Animal Cognition*, *22*, 445–452. https://doi.org/10.1007/s10071-019-01250-9

Ferris, C. F., Axelson, J. F., Shinto, L. H., & Albers, H. E. (1987). Scent marking and the maintenance of dominant/subordinate status in male golden hamsters. *Physiology and Behavior*, *40*(5), 661–664. https://doi.org/10.1016/0031-9384(87)90114-4

Finger, J. H., Smith, C. M., Hayamizu, T. F., McCright, I. J., Xu, J., Law, M., … Blodgett, O. (2016). The mouse gene expression database (GXD): 2017 update. *Nucleic Acid Research*, *45*(D1), D730–D736.

Freeman, A. R., Sheehan, M. J., & Ophir, A. G. (2019). Anogenital distance predicts sexual odour preference in African giant pouched rats. *Animal Behaviour*, *148*, 123–132. https://doi.org/10.1016/j.anbehav.2018.12.010

Ganem, G., Rueff, C., & Perriat-Sanguinet, M. (2014). The genetic architecture of chemosensory cues involved in species recognition: A behavioral approach in the house mouse. *Behavior Genetics*, *44*(1), 56–67. https://doi.org/10.1007/s10519-013-9621-y

Geraldes, A., Basset, P., Gibson, B., Smith, K. L., Harr, B., Yu, H.-T., … Nachman, M. W. (2008). Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology*, *17*(24), 5349–5363. https://doi.org/10.1111/j.1365-294X.2008.04005.x

Gomez-Baena, G., Armstrong, S. D., Halstead, J. O., Prescott, M., Roberts, S. A., McLean, L., … Beynon, R. (2018). Molecular complexity of the major urinary protein system of the Norway rat, *Rattus norvegicus*. *bioRxiv*, 478362.

Gosling, L. M., & Roberts, S. C. (2001). Scent-marking by male mammals: Cheat-proof signals to competitors and mates. *Advances in the study of behavior* (Vol. *30*, pp. 169–217). San Diego, CA: Academic Press.

Green, J. P., Holmes, A. M., Davidson, A. J., Paterson, S., Stockley, P., Beynon, R. J., & Hurst, J. L. (2015). The genetic basis of kin recognition in a cooperatively breeding mammal. *Current Biology*, *25*, 2631–2641. https://doi.org/10.1016/j.cub.2015.08.045

Hagemeyer, P., Begall, S., Janotova, K., Todrank, J., Heth, G., Jedelsky, P. L., … Stopka, P. (2011). Searching for major urinary proteins (MUPs) as chemosignals in urine of subterranean rodents. *Journal of Chemical Ecology*, *37*(7), 687–694. https://doi.org/10.1007/s10886-011-9971-y

Hoffman, E., Pickavance, L., Thippeswamy, T., Beynon, R. J., & Hurst, J. L. (2015). The male sex pheromone darcin stimulates hippocampal neurogenesis and cell proliferation in the subventricular zone in female mice. *Frontiers in Behavioral Neuroscience*, *9*, 106. https://doi.org/10.3389/fnbeh.2015.00106

Hurst, J. L. (1987). The functions of urine marking in a free-living population of house mice, *Mus domesticus* Rutty. *Animal Behaviour*, *35*, 1433–1442. https://doi.org/10.1016/S0003-3472(87)80016-7

Hurst, J. L. (2009). Female recognition and assessment of males through scent. *Behavioural Brain Research*, *200*, 295–303. https://doi.org/10.1016/j.bbr.2008.12.020

Hurst, J. L., & Beynon, R. J. (2004). Scent wars: The chemobiology of competitive signalling in mice. *BioEssays*, *26*, 1288–1298. https://doi.org/10.1002/bles.20147

Hurst, J. L., & Beynon, R. J. (2013). Rodent urinary proteins: genetic identity signals and pheromones. In M. L. East & M. Dehnhard (Eds.), *Chemical signals in vertebrates 12* (pp. 117–133). Berlin, Germany: Springer.

Hurst, J. L., Beynon, R. J., Armstrong, S. D., Davidson, A. J., Roberts, S. A., Gómez-Baena, G., … Ganem, G. (2017). Molecular heterogeneity in major urinary proteins of *Mus musculus* subspecies: Potential candidates involved in speciation. *Scientific Reports*, *7*, 44992. https://doi.org/10.1038/srep44992

Hurst, J. L., Payne, C. E., Nevison, C. M., Marie, A. D., Humphries, R. E., Robertson, D. H. L., … Beynon, R. J. (2001). Individual recognition in mice mediated by major urinary proteins. *Nature*, *414*, 631–634. https://doi.org/10.1038/414631a

James, T. Y. (2012). Ancient yet fast: Rapid evolution of mating genes and mating systems in fungi. In R. S. Singh, J. Xu, & R. J. Kulathinal (Eds.), *Rapidly evolving genes and genetic systems*. Oxford, UK: Oxford University Press.

Janssenswillen, S., Vandebergh, W., Treer, D., Willaert, B., Maex, M., Van Bocxlaer, I., & Bossuyt, F. (2014). Origin and diversification of a salamander sex pheromone system. *Molecular Biology and Evolution*, *32*(2), 472–480. https://doi.org/10.1093/molbev/msu316

Johnston, R. E. (2003). Chemical communication in rodents: From pheromones to individual recognition. *Journal of Mammalogy*, *84*(4), 1141–1162. https://doi.org/10.1644/BLe-010

Johnstone, R. A. (1997). Recognition and the evolution of distinctive signatures: When does it pay to reveal identity? *Proceedings of the Royal Society B: Biological Sciences*, *264*(1387), 1547–1553. https://doi.org/10.1098/rspb.1997.0215

Kaur, A. W., Ackels, T., Kuo, T.-H., Cichy, A., Dey, S., Hays, C., … Stowers, L. (2014). Murine pheromone proteins constitute a context-dependent

combinatorial code governing multiple social behaviors. *Cell*, *157*(3), 676–688. https://doi.org/10.1016/j.cell.2014.02.025

Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., … Adams, D. J. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, *477*(7364), 289. https://doi.org/10.1038/nature10413

Lassance, J.-M., Groot, A. T., Liénard, M. A., Antony, B., Borgwardt, C., Andersson, F., … Löfstedt, C. (2010). Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. *Nature*, *466*(7305), 486. https://doi.org/10.1038/nature09058

Lassance, J.-M., Lienard, M. A., Antony, B., Qian, S., Fujii, T., Tabata, J., … Lofstedt, C. (2013). Functional consequences of sequence variation in the pheromone biosynthetic gene pgFAR for Ostrinia moths. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(10), 3967–3972. https://doi.org/10.1073/pnas.1208706110

Logan, D. W., Marton, T. F., & Stowers, L. (2008). Species specificity in major urinary proteins by parallel evolution. *PLoS ONE*, *3*, e3280. https://doi.org/10.1371/journal.pone.0003280

Maddison, W. P. (2008). MESQUITE: A modular system for evolutionary analysis. *Evolution*, *62*, 1103–1118.

Martín, J., & Lopez, P. (2007). Scent may signal fighting ability in male Iberian rock lizards. *Biology Letters*, *3*(2), 125–127. https://doi.org/10.1098/rsbl.2006.0589

Martín-Sánchez, A., McLean, L., Beynon, R. J., Hurst, J. L., Ayala, G., Lanuza, E., & Martinez-Garcia, F. (2015). From sexual attraction to maternal aggression: When pheromones change their behavioural significance. *Hormones and Behavior*, *68*, 65–76. https://doi.org/10.1016/j.yhbeh.2014.08.007

Mateo, J. M. (2003). Kin recognition in ground squirrels and other rodents. *Journal of Mammalogy*, *84*(4), 1163–1181. https://doi.org/10.1644/BLe-011

Mudge, J. M., Armstrong, S. D., McLaren, K., Beynon, R. J., Hurst, J. L., Nicholson, C., … Harrow, J. L. (2008). Dynamic instability of the major urinary protein gene family revealed by genomic and phenotypic comparisons between C57 and 129 strain mice. *Genome Biology*, *9*(5), R91. https://doi.org/10.1186/gb-2008-9-5-r91

Mullen, S. P., Mendelson, T. C., Schal, C., & Shaw, K. L. (2007). Rapid evolution of cuticular hydrocarbons in a species radiation of acoustically diverse Hawaiian crickets (Gryllidae: Trigonidiinae: Laupala). *Evolution*, *61*(1), 223–231. https://doi.org/10.1111/j.1558-5646.2007.00019.x

Nelson, A. C., Cunningham, C. B., Ruff, J. S., & Potts, W. K. (2015). Protein pheromone expression levels predict and respond to the formation of social dominance networks. *Journal of Evolutionary Biology*, *28*(6), 1213–1224. https://doi.org/10.1111/jeb.12643

Nevison, C. M., Armstrong, S., Beynon, R. J., Humphries, R. E., & Hurst, J. L. (2003). The ownership signature in mouse scent marks is involatile. *Proceedings of the Royal Society B: Biological Sciences*, *270*, 1957–1963. https://doi.org/10.1098/rspb.2003.2452

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274.

Phelan, M. M., McLean, L., Hurst, J. L., Beynon, R. J., & Lian, L.-Y. (2014). Comparative study of the molecular variation between 'central' and 'peripheral' MUPs and significance for behavioural signalling. *Biochemical Society Transactions*, *42*(4), 866–872. https://doi.org/10.1042/BST20140082

Phifer-Rixey, M., & Nachman, M. W. (2015). The natural history of model organisms: Insights into mammalian biology from the wild house mouse *Mus musculus*. *eLife*, *4*, e05959. https://doi.org/10.7554/eLife.05959

Pocock, M. J. O., Searle, J. B., & White, P. C. L. (2004). Adaptations of animals to commensal habitats: Population dynamics of house mice Mus musculus domesticus on farms. *Journal of Animal Ecology*, *73*, 878–888. https://doi.org/10.1111/j.0021-8790.2004.00863.x

Pollard, K. A., & Blumstein, D. T. (2011). Social group size predicts the evolution of individuality. *Current Biology*, *21*, 413–417. https://doi.org/10.1016/j.cub.2011.01.051

Ramm, S. A., Cheetham, S. A., & Hurst, J. L. (2008). Encoding choosiness: Female attraction requires prior physical contact with individual male scents in mice. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 1727–1735. https://doi.org/10.1098/rspb.2008.0302

Roberts, S. A., Davidson, A. J., McLean, L., Beynon, R. J., & Hurst, J. L. (2012). Pheromonal induction of spatial learning in mice. *Science*, *338*(6113), 1462–1465.

Roberts, S. A., Prescott, M. C., Davidson, A. J., McLean, L., Beynon, R. J., & Hurst, J. L. (2018). Individual odour signatures that mice learn are shaped by involatile major urinary proteins (MUPs). *BMC Biology*, *16*(1), 48. https://doi.org/10.1186/s12915-018-0512-9

Roberts, S. A., Simpson, D. M., Armstrong, S. D., Davidson, A. J., Robertson, D. H., McLean, L., … Hurst, J. L. (2010). Darcin: A male pheromone that stimulates female memory and sexual attraction to an individual male's odour. *BMS Biology*, *8*, https://doi.org/10.1186/1741-7007-8-75

Robertson, D. H. L., Hurst, J. L., Searle, J. B., Gunduz, I., & Beynon, R. J. (2007). Characterization and comparison of major urinary proteins from the house mouse, *Mus musculus domesticus*, and the aboriginal mouse, *Mus macedonicus*. *Journal of Chemical Ecology*, *33*, 613–630. https://doi.org/10.1007/s10886-006-9247-0

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24. https://doi.org/10.1038/nbt.1754

Seeholzer, L. F., Seppo, M., Stern, D. L., & Ruta, V. (2018). Evolution of a central neural circuit underlies Drosophila mate preferences. *Nature*, *559*(7715), 564. https://doi.org/10.1038/s41586-018-0322-9

Sheehan, M. J., Lee, V., Corbett-Detig, R., Bi, K., Beynon, R. J., Hurst, J. L., & Nachman, M. W. (2016). Selection on coding and regulatory variation maintains individuality in major urinary protein scent marks in wild mice. *PLoS Genetics*, *12*(3), e1005891. https://doi.org/10.1371/journal.pgen.1005891

Sheehan, M. J., Miller, C., & Reeve, H. K. (2017). Identity signaling and patterns of cooperative behavior. *Integrative and Comparative Biology*, *57*(3), 580–588. https://doi.org/10.1093/icb/icx054

Sheehan, M. J., & Tibbetts, E. A. (2009). Evolution of identity signals: Frequency-dependent benefits of distinctive phenotypes used for individual recognition. *Evolution*, *63*(12), 3106–3113. https://doi.org/10.1111/j.1558-5646.2009.00833.x

Sinervo, B., Bleay, C., & Adamopoulou, C. (2001). Social causes of correlational selection and the resolution of a heritable throat color polymorphism in a lizard. *Evolution*, *55*, 2040–2052. https://doi.org/10.1111/j.0014-3820.2001.tb01320.x

Smadja, C., Catalan, J., & Ganem, G. (2004). Strong premating divergence in a unimodal hybrid zone between two subspecies of the house mouse. *Journal of Evolutionary Biology*, *17*(1), 165–176. https://doi.org/10.1046/j.1420-9101.2003.00647.x

Smadja, C., & Ganem, G. (2002). Subspecies recognition in the house mouse: A study of two populations from the border of a hybrid zone. *Behavioral Ecology*, *13*(3), 312–320.

Steppan, S. J., & Schenk, J. J. (2017). Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. *PLoS ONE*, *12*(8), e0183070. https://doi.org/10.1371/journal.pone.0183070

Stopková, R., Stopka, P., Janotova, K., & Jedelsky, P. L. (2007). Species-specific expression of major urinary proteins in the house mice (*Mus musculus musculus* and *Mus musculus domesticus*). *Journal of Chemical Ecology*, *33*, 861–869. https://doi.org/10.1007/s10886-007-9262-9

Stopková, R., Vinkler, D., Kuntová, B., Šedo, O., Albrecht, T., Suchan, J., … Stopka, P. (2016). Mouse lipocalins (MUP, OBP, LCN) are co-expressed

in tissues involved in chemical communication. *Frontiers in Ecology and Evolution*, *4*, 47. https://doi.org/10.3389/fevo.2016.00047

Symonds, M. R., & Elgar, M. A. (2008). The evolution of pheromone diversity. *Trends in Ecology and Evolution*, *23*(4), 220–228. https://doi.org/10.1016/j.tree.2007.11.009

Thom, M. D., & Hurst, J. L. (2004). Individual recognition by scent. *Annales Zoologici Fennici*, *41*, 765–787.

Thom, M. D., Stockley, P., Jury, F., Ollier, W. E. R., Beynon, R. J., & Hurst, J. L. (2008). The direct assessment of genetic heterozygosity through scent in the mouse. *Current Biology*, *18*, 619–623. https://doi.org/10.1016/j.cub.2008.03.056

Thonhauser, K. E., Raveh, S., Hettyey, A., Beissmann, H., & Penn, D. J. (2013). Scent marking increases male reproductive success in wild house mice. *Animal Behavior*, *86*(5), 1013–1021. https://doi.org/10.1016/j.anbehav.2013.09.004

Thoß, M., Enk, V., Yu, H., Miller, I., Luzynski, K. C., Balint, B., … Penn, D. J. (2016). Diversity of major urinary proteins (MUPs) in wild house mice. *Scientific Reports*, *6*, 38378. https://doi.org/10.1038/srep38378

Thoß, M., Luzynski, K., Ante, M., Miller, I., & Penn, D. J. (2015). Major urinary protein (MUP) profiles show dynamic changes rather than individual 'barcode' signatures. *Frontiers in Ecology and Evolution*, *3*, 71. https://doi.org/10.3389/fevo.2015.00071

Thoß, M., Luzynski, K. C., Enk, V. M., Razzazi-Fazeli, E., Kwak, J., Ortner, I., & Penn, D. J. (2019). Regulation of volatile and non-volatile pheromone attractants depends upon male social status. *Scientific Reports*, *9*(1), 489. https://doi.org/10.1038/s41598-018-36887-y

Thünken, T., Waltschyk, N., Bakker, T. C., & Kullmann, H. (2009). Olfactory self-recognition in a cichlid fish. *Animal Cognition*, *12*(5), 717–724. https://doi.org/10.1007/s10071-009-0231-2

Thybert, D., Roller, M., Navarro, F. C., Fiddes, I., Streeter, I., Feig, C., … Akanni, W. (2018). Repeat associated mechanisms of genome evolution and function revealed by the *Mus caroli* and *Mus pahari* genomes. *Genome Research*, *28*(4), 448–459.

Tibbetts, E. A., Mullen, S. P., & Dale, J. (2017). Signal function drives phenotypic and genetic diversity: The effects of signalling individual identity, quality or behavioural strategy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1724), 20160347. https://doi.org/10.1098/rstb.2016.0347

Timm, D. E., Baker, L. J., Mueller, H., Zidek, L., & Novotny, M. V. (2001). Structural basis of pheromone binding to mouse major urinary protein (MUP-I). *Protein Science*, *10*(5), 997–1004. https://doi.org/10.1110/ps.52201

Todrank, J., & Heth, G. (2003). Odor–genes covariance and genetic relatedness assessments: Rethinking odor-based "recognition" mechanisms in rodents. *Advances in the Study of Behavior*, *32*, 77–130.

Tupec, M., Buček, A., Janoušek, V., Vogel, H., Prchalová, D., Kindl, J., … Pichová, I. (2019). Expansion of the fatty acyl reductase gene family shaped pheromone communication in Hymenoptera. *eLife*, *8*, e39231. https://doi.org/10.7554/eLife.39231

Willse, A., Kwak, J., Yamazaki, K., Preti, G., Wahl, J. H., & Beauchamp, G. K. (2006). Individual odortypes: Interaction of MHC and background genes. *Immunogenetics*, *58*(12), 967–982. https://doi.org/10.1007/s00251-006-0162-x

Wyatt, T. D. (2010). Pheromones and signature mixtures: Defining species-wide signals and variable cues for identity in both invertebrates and vertebrates. *Journal of Comparative Physiology*, *196*(10), 685–700. https://doi.org/10.1007/s00359-010-0564-y

Wyatt, T. D. (Ed.) (2014). *Pheromones and animal behavior: Chemical signals and signatures*. Cambridge, UK: Cambridge University Press.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Sheehan MJ, Campbell P, Miller CH. Evolutionary patterns of major urinary protein scent signals in house mice and relatives. *Mol Ecol*. 2019;00:1–15. https://doi.org/10.1111/mec.15155