



**HAL**  
open science

## Do changes in gene expression contribute to sexual isolation and reinforcement in the house mouse?

Etienne Loire, Sergio Tusso, Pierre Caminade, Dany Severac, Pierre Boursot, Guila Ganem, Carole Smadja

### ► To cite this version:

Etienne Loire, Sergio Tusso, Pierre Caminade, Dany Severac, Pierre Boursot, et al.. Do changes in gene expression contribute to sexual isolation and reinforcement in the house mouse?. *Molecular Ecology*, 2017, 26 (19), pp.5189 - 5202. 10.1111/mec.14212 . hal-01815471

**HAL Id: hal-01815471**

**<https://hal.umontpellier.fr/hal-01815471>**

Submitted on 5 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Title

2 Do changes in gene expression contribute to sexual isolation and  
3 reinforcement in the house mouse?

4

5 Etienne Loire <sup>§,1,2</sup>, Sergio Tusso <sup>§,1,4</sup>, Pierre Caminade <sup>1</sup>, Dany Severac <sup>3</sup>, Pierre Boursot <sup>1</sup>, Guila  
6 Ganem <sup>1</sup> and Carole M. Smadja <sup>1</sup>

7

8 <sup>§</sup>These authors contributed equally

9 <sup>1</sup>Institut des Sciences de l'Evolution UMR 5554 (Centre National de la Recherche Scientifique  
10 CNRS, Institut pour la Recherche et le Développement IRD, EPHE, Université de Montpellier),  
11 cc065 Université de Montpellier, 34095 Montpellier, France

12 <sup>2</sup> Current address: CIRAD, UMR15 CMAEE, Campus international de Baillarguet, 34398  
13 Montpellier Cedex 5, France

14 <sup>3</sup> MGX-Montpellier GenomiX; c/o Institut de Génomique Fonctionnelle; 141 rue de la  
15 cardonille; 34094 Montpellier Cedex 5; France

16 <sup>4</sup> Current address: Evolutionsbiologiskt Centrum, Norbyvägen 18D, University of Uppsala,  
17 75236 Uppsala, Sweden

18

19 **Corresponding author:** Carole M. Smadja: [carole.smadja@univ-montp2.fr](mailto:carole.smadja@univ-montp2.fr)

20

21 **Running title:** candidate genes for sexual isolation in mice

22

23 **Keywords:** assortative mate preference, speciation, hybrid zone, RNA-seq, olfaction,  
24 vomeronasal receptors

25

26 Abstract

27 Expression divergence, rather than sequence divergence, has been shown to be important in  
28 speciation, particularly in the early stages of divergence of traits involved in reproductive  
29 isolation. In the two European subspecies of house mice, *Mus musculus musculus* and *M. m.*  
30 *domesticus*, earlier studies have demonstrated olfactory-based assortative mate preference  
31 in populations close to their hybrid zone. It has been suggested that this behaviour evolved  
32 following the recent secondary contact between the two taxa (~3,000 years ago) in response  
33 to selection against hybridisation. To test for a role of changes in gene expression in the  
34 observed behavioural shift, we conducted a RNA-sequencing experiment on mouse  
35 vomeronasal organs. Key candidate genes for pheromone-based subspecies recognition, the  
36 vomeronasal receptors, are expressed in these organs. Overall patterns of gene expression  
37 varied significantly between samples from the two subspecies, with a large number of  
38 differentially expressed genes between the two taxa. In contrast, only ~200 genes were  
39 found repeatedly differentially expressed between populations within *M. m. musculus* that  
40 did, or did not display assortative mate preferences (close to, or more distant from the  
41 hybrid zone, respectively), with an overrepresentation of genes belonging to vomeronasal  
42 receptor family 2. These receptors are known to play a key role in recognition of chemical  
43 cues that handle information about genetic identity. Interestingly, four out of five of these  
44 differentially expressed receptors belong to the same phylogenetic cluster, suggesting  
45 specialisation of a group of closely-related receptors in the recognition of odorant signals  
46 that may allow subspecies recognition and assortative mating.

47

48

## 49 **Introduction**

50 Understanding the genetic basis of speciation remains a key challenge in evolutionary  
51 biology (Seehausen *et al.* 2014), in particular determining the genetic architecture of  
52 differentiation and reproductive isolation (Nosil & Feder 2012), the types of genes involved  
53 (Nosil & Schluter, 2011), and the types of genetic changes responsible for the evolution of  
54 reproductive isolation between diverging taxa (Hoekstra & Coyne 2007). Although evidence  
55 is rapidly accumulating for the role of protein coding and structural changes in adaptation  
56 and reproductive isolation (Stapley *et al.* 2010), it has also long been argued that gene  
57 expression differences may be of prime importance in species differentiation (King & Wilson  
58 1975; Carroll *et al.* 2005), allowing rapid divergence (Wolf *et al.* 2010). Moreover, measures  
59 of gene expression provide an important bridge between genotype and phenotype (Huestis  
60 & Marshall 2009). In recent years, the study of gene expression has been greatly facilitated  
61 by high-throughput sequencing-based methods such as RNA-seq (Mortazavi *et al.* 2008;  
62 Wang *et al.* 2009), contributing to understand the importance of expression variation in both  
63 local adaptation and speciation.

64

### 65 *Gene expression changes and the evolution of reproductive isolation*

66 There are many examples of differential gene expression causing adaptive phenotypic  
67 changes (e.g., Chan *et al.* 2010a; McBride *et al.* 2014), and large-scale transcriptomics  
68 studies have shown heritable expression differences between diverging populations (e.g.,  
69 Wolf *et al.* 2010; Uebbing *et al.* 2016; Davidson & Balakrishnan, 2016). How these  
70 components of gene expression can be associated with reproductive isolation is still a largely  
71 unexplored question. The major advances on this point come from studies investigating the  
72 role of gene regulation in the evolution of postzygotic isolation and, in particular, how  
73 misexpression in hybrids can reflect Dobzhansky-Muller incompatibilities (e.g. Haerty &  
74 Singh, 2006; Ortiz-Barrientos, Counterman, & Noor, 2007; Dion-Cote, Renaut, Normandeau,

75 & Bernatchez, 2014; Davidson & Balakrishnan, 2016; Mack *et al.* 2016). A recent study also  
76 identified significant changes in expression associated with a postmating-prezygotic  
77 reproductive barrier between two *Solanum* plant species, that involve effector molecules  
78 possibly responsible for pollen tube arrest (Pease *et al.* 2016). As far as premating isolation is  
79 concerned, evidence is mostly found in phytophagous insect species that are reproductively  
80 isolated due to host plant preference and specialisation. For example, recent studies have  
81 shown expression differences between populations specialised on different plants, involving  
82 genes related to host selection (odorant receptors, salivary proteins, detoxification genes)  
83 (e.g. cactus-specialised populations of *Drosophila mettleri*: Hoang, Matzkin, & Bono, 2015;  
84 *Rhagoletis* host races: Ragland *et al.* 2015; pea aphid host races: Eyres *et al.* 2016). These  
85 differences in expression have therefore been proposed to play a key role for adaptive  
86 genetic divergence in traits causing reproductive isolation during ecological speciation  
87 (Pavey, Collin, Nosil, & Rogers, 2010; Etges, 2014). Despite growing evidence for the role of  
88 gene regulation in mate choice and sexual selection (Wilkinson *et al.* 2015), support for  
89 expression changes contributing to sexual isolation is relatively scarce, and again concerns  
90 insect species. Examples include sex-specific gene expression changes underlying shifts in  
91 sex-pheromone preferences between populations of the mosquito *Anopheles gambiae*  
92 (Cassone *et al.* 2008), the silkworm *Bombyx mori* (Fujii *et al.* 2011) and *D. melanogaster*  
93 (Bailey *et al.* 2011), and a change in expression of a fatty acid synthase responsible for  
94 cuticular hydrocarbon production inducing sexual isolation between *D. serrata* and *D. birchii*  
95 (Chung *et al.* 2014).

96

97 *Gene expression and the evolution of assortative mate preference in a house mouse hybrid*  
98 *zone*

99 In this study, we investigate the role of expression divergence in the evolution of olfactory-  
100 based assortative mate preference between two subspecies of the house mouse (*Mus*

101 *musculus*) in order to gain insights into the genetic basis of sexual isolation in this  
102 mammalian system. The two European subspecies of the house mouse, *M. m. musculus* and  
103 *M. m. domesticus* have diverged in allopatry for 0.5 million years in the Indian sub-continent  
104 (Boursot *et al.* 1996), before meeting secondarily in Europe around 5,000 years ago (and  
105 around 3,000 years ago in Denmark) (Cucchi *et al.* 2012). They now form a hybrid zone  
106 stretching from Norway to the Black Sea. Assortative mate preference between the two  
107 subspecies has been demonstrated in populations at the edges of the hybrid zone in  
108 Denmark in both subspecies, and in both males and females, but appears particularly strong  
109 in *M. m. musculus* populations (Smadja & Ganem 2002, 2005; Smadja *et al.* 2004a; Ganem *et*  
110 *al.* 2008). These results suggest a degree of sexual isolation between the two taxa. In  
111 contrast, populations of both subspecies occurring in distant allopatry or slightly further  
112 away from the Danish hybrid zone do not display any directional mate preference (Smadja &  
113 Ganem 2005; Smadja *et al.* 2015). This pattern of reproductive character displacement has  
114 been confirmed in the Czech part of the hybrid zone (Bimova *et al.* 2011). Since selection  
115 against hybrids is acting in the mouse hybrid zone (reduced hybrid fertility: Britton-Davidian  
116 *et al.* 2005; Albrechtova *et al.* 2012; Turner *et al.* 2012; Bhattacharyya *et al.* 2013; Turner &  
117 Harr 2014; sexual selection against hybrids: Latour *et al.* 2014; hybrid microbiome  
118 dysfunction: Wang *et al.* 2015) a reinforcement process has been hypothesised in this  
119 system, by which premating barriers could have evolved as a response to selection against  
120 hybrids (Bimova *et al.* 2011; Smadja & Ganem, 2005; Smadja *et al.* 2015).

121

122 A current challenge is to determine the genetic basis of assortative mate preferences  
123 between the two subspecies and of the behavioural shift within each subspecies between  
124 populations distant from or close to the hybrid zone. Previous studies have shown that  
125 chemical signals present in urine are sufficient to elicit discrimination and assortative mate  
126 preferences between the two subspecies (Smadja & Ganem 2002, 2005; Smadja *et al.*

127 2004a; Ganem *et al.* 2008; Bimova *et al.* 2011), suggesting a role of the olfactory system in  
128 this behavioural shift. In mammals, two organs are responsible for odorant signal  
129 recognition: the main olfactory epithelium (MOE), primarily involved in recognition of  
130 airborne chemicals present in the environment (odorants and pheromones), and the  
131 accessory olfactory organ or vomeronasal organ (VNO), known to play a key role in both  
132 volatile and non-volatile kairomone and pheromone recognition (Isogai *et al.* 2011;  
133 Chamero, Leinders-Zufall, & Zufall, 2012; Liberles, 2014; Bear *et al.* 2016). Recognition in  
134 these organs involves chemoreceptors (olfactory receptors (ORs) in the MOE and  
135 vomeronasal receptors (VRs) in the VNO) that are organised in very large multigene families  
136 in the mouse genome (more than 1,200 ORs and 530 VRs, Ibarra-Soria, Levitin, & Logan,  
137 2013; Ibarra-Soria, Levitin, Saraiva, & Logan, 2014; Zhang & Firestein, 2002). Although  
138 duplications, mutations, and adaptive evolution certainly play an important role in the  
139 evolution of chemoreceptors and their associated behaviours (Horth, 2007; Karn *et al.* 2010;  
140 Smadja *et al.* 2015), variation in expression at these genes can also strongly influence mouse  
141 behaviour (Keller *et al.* 2009; Zhang *et al.* 2010; Bear *et al.* 2016). A recent study has  
142 demonstrated that changes in expression patterns of vomeronasal receptors can induce  
143 changes in patterns of mate preference between laboratory strains of the house mouse  
144 (Broad & Keverne 2012), confirming the key role of the vomeronasal organ and its receptors  
145 in the recognition and transduction of olfactory signals used in mate choice, as well as the  
146 possible link between variation in VR expression and variation in mate preference in mice.

147

#### 148 *Predictions and experimental design*

149 We hypothesized that the observed behavioural divergence between populations close to  
150 the hybrid zone and displaying strong assortative mate preferences (hereafter 'Choosy'  
151 populations) and populations further away from the hybrid zone and not displaying any  
152 directional mate preferences (hereafter 'Non-Choosy' populations) could originate from

153 divergence in expression of some vomeronasal receptors. To test this hypothesis, we  
154 conducted deep RNA-sequencing of whole vomeronasal organs, an approach that has  
155 already been validated for detection of expression variation at VRs (Ibarra-Soria *et al.* 2014).  
156 We compared adult female and male mice originating from different 'Non-Choosy' and  
157 'Choosy' populations of *M. m. musculus* and *M. m. domesticus*. Estimates of differential  
158 expression allowed us to identify specific sets of vomeronasal receptors and quantify  
159 differences in receptor expression potentially related to the shift toward assortative mate  
160 preference in the hybrid zone.

161



162 **MATERIAL AND METHODS**

163

164 **Samples**

165 *Mice used in the study*

166 Patterns of gene expression were studied in adult mice that were born in the wild, trapped  
167 in Jutland, Denmark in October 2010 in several sites (indoor farms and other human  
168 dwellings) and then maintained in the laboratory under controlled conditions before being  
169 behaviourally tested and euthanised for dissection. As it is known that postnatal odour  
170 exposure can influence patterns of mate choice and VNO gene expression of adult mice  
171 (Broad & Keverne 2012; Chamero *et al.* 2012; Cadiou *et al.* 2014), this experimental strategy  
172 allowed us to study behavioural and expression patterns characterising wild animals, while  
173 controlling for some factors (odorant environment, age range, diet) just before dissection  
174 that could inflate variance in expression among samples.

175

176 Mice were trapped in three distinct sampling areas, characterised by populations with  
177 distinct mate preference behaviours (Figure S1): (1) the border of the hybrid zone on the *M.*  
178 *musculus* side (50 km North to the genetic centre of hybrid zone (defined in Raufaste *et*  
179 *al.* 2005)) where strong assortative mating has been documented (*'musculus Choosy'*  
180 samples) (Smadja *et al.* 2004b, 2015; Smadja & Ganem 2005; Ganem *et al.* 2008), (2)  
181 another area in Denmark but further away from the hybrid zone, where mate preference  
182 was not assortative (*'musculus Non-Choosy'* samples) (Smadja *et al.* 2015); and (3) the  
183 border of the hybrid zone on the *M. m. domesticus* side (50 km South to the genetic centre  
184 of the hybrid zone), where weak assortative mating was documented (*'domesticus'* Choosy  
185 samples) (Smadja & Ganem 2005; Ganem *et al.* 2008; Latour *et al.* 2014). The *M. m.*  
186 *domesticus* samples were used as a control for the mapping of *M. m. musculus* reads against  
187 the house mouse reference genome, which is primarily of *M. m. domesticus* origin, and for

188 interspecific comparisons. We decided to focus on the *M. m. musculus* side for comparisons  
189 between ‘Choosy’ and ‘Non-Choosy’ populations, because the strongest behavioural  
190 contrasts were observed there (Smadja *et al.* 2004b; Smadja & Ganem 2005; Ganem *et al.*  
191 2008). Two populations per geographical area were included as biological replicates, each  
192 composed of several trapping sites (six populations in total).

193

194 Mice were brought back to the laboratory in Montpellier (France) and maintained in  
195 quarantine under controlled and homogenized conditions for two months (same facility  
196 room, odorant environment, 12/12 photoperiod, and diet - food ad libitum). Only mice  
197 trapped as young adults were selected for this study, so that the ages of individuals  
198 participating to the behavioural and expression analyses were likely to have been within the  
199 age range at which VR expression in the VNO is optimal (between 90 and 120 days old,  
200 Zhang *et al.* 2010). During the period in the laboratory, mice were tested for sex odour  
201 preferences to confirm assortative preferences in the ‘*musculus* Choosy’ group and an  
202 absence of directional preference in the ‘*musculus* Non Choosy’ group (protocol and results  
203 reported in Smadja *et al.* 2015). We also confirmed weak assortative mate preferences in  
204 the ‘*domesticus* Choosy’ group (Latour *et al.* 2014). We then selected eight males and eight  
205 females per population, resulting in a total of 96 adult male and female mice taking part to  
206 the RNA-seq experiment. The average mate preference of RNA-seq samples did not deviate  
207 from the average mate preference assessed in larger population samples. Geographical,  
208 genetic and behavioural characteristics of the mice used in this study are summarised in  
209 Table S1 (the method used to estimate hybridisation index of each mouse is described in  
210 detail in Latour *et al.* 2014 and Smadja *et al.* 2015 and summarised in Table S1).

211

212 *Vomeronasal organ dissection and RNA extraction*

213 For each of the 96 female and male mice, the VNO was dissected rapidly after death by  
214 cervical dislocation. The VNO was immediately immersed in RNAlater and stored at -80C.  
215 Total RNA was extracted using Macherey-Nagel XS kits or the phenol-chloroform based  
216 TRIzol procedure. RNA extracts were subsequently quantified and checked for quality with  
217 Nanodrop to assess sample concentration and purity and with Agilent Bioanalyzer to assess  
218 sample integrity. To constitute samples representative of each of the six populations and  
219 sexes, and to obtain enough material for library preparation, equimolar mixes from the eight  
220 males or eight females per population were pooled and again checked for quantity and  
221 quality using the Agilent Bioanalyzer and Nanodrop, to obtain 3 to 5 µg per pool (200ng/µL).

222

223 Trapping and transport of mice were performed under permission from the French Ministry  
224 of Agriculture (authorization n°C34-265). Housing took place in the facilities of the University  
225 of Montpellier and followed the ethical and animal welfare guidelines of the University of  
226 Montpellier and the regional ethical committee.

227

## 228 **Transcriptome sequencing and raw data analysis**

### 229 *Library preparation and sequencing*

230 RNA-seq libraries were prepared using the TruSeq RNA sample preparation kit (Illumina) and  
231 quantified using Agilent DNA 1000 Labchip kits (Agilent Technologies, Palo Alto, CA). The  
232 libraries were then sequenced on an Illumina HiSeq 2000 platform using a paired-end 100bp  
233 read-length protocol (3 libraries per lane, 4 lanes in total). Library preparation and  
234 sequencing were carried out at the MGX -Montpellier Genomix platform (Institut de  
235 Génomique Fonctionnelle- Institut de Génétique Humaine, Montpellier, France).

236

### 237 *Filtering and mapping*

238 Raw reads were filtered in order to discard reads with low quality sequence. The remaining

239 reads were trimmed to subtract primer sequences and low quality 3' ends using *cutadapt*  
240 (Martin 2011). The package *RSEM* (Li & Dewey 2011), with default mapping parameters, was  
241 used to map the reads against the house mouse reference genome and its known transcript  
242 annotation database (<http://www.ensembl.org/>, version GRCm38.p2) and to estimate gene  
243 expression abundance. Since the reference genome used for mapping corresponds to an  
244 inbred strain of the laboratory mouse (C57BL/6J) (primarily of *M. m. domesticus* origin), we  
245 compared the distribution of expression estimations between wild *M. m. domesticus* and *M.*  
246 *m. musculus* samples to detect any potential mapping biases among samples.

247

#### 248 *Identification of Vomeronasal Receptor and Olfactory Receptor transcripts*

249 Coding sequences belonging to family 1 (V1R) and family 2 (V2R) of vomeronasal receptors  
250 (Yang & Zhang 2007; Zhang *et al.* 2007) as well as coding sequences of ORs were extracted  
251 with the BioMart tool from Ensembl (<http://www.ensembl.org/>). Information on  
252 chromosome location, genomic coordinates and strand for each gene was according to the  
253 NCBI-Build-37 mouse assembly (mm9).

254

#### 255 **Expression analyses**

256 To compare the amount of gene expression among different classes of genes (e.g. VR genes  
257 versus non-VR genes), we normalised coverage values using the Bioconductor package *CQN*  
258 (conditional quantile normalization) (version 1.2.0) (Hansen *et al.* 2012), which accounts for  
259 possible bias due to library size, gene length and GC content. For comparisons of gene  
260 expression patterns among samples and differential expression analyses, read counts were  
261 normalised using the Bioconductor package *edgeR* (version 2.13) (McCarthy *et al.* 2012)  
262 implemented for the software *R* v.3.0.1 (<http://www.r-project.org/>) and the method of  
263 trimmed mean of M values (TMM) (Robinson *et al.* 2010), which normalises for the library  
264 size and for the total number of expressed transcripts in each sample. A gene was

265 considered as expressed in one sample if the number of reads was higher than 1 count per  
266 million (CPM).

267

268 Multidimensional scaling (MDS) plots were generated from the 500 genes maximizing the  
269 distances between points, to visualise similarity of global gene expression between samples.

270 Coverage at the gene level was obtained by summing up transcript coverage values

271 associated with a given gene. To assess whether global patterns of gene expression vary

272 between subspecies, mate preference profile groups ('*musculus* Choosy' versus '*musculus*

273 Non-Choosy') and sexes, we first verified equal dispersions among samples and then

274 performed permutation-based MANOVA analyses (Anderson 2001) using the PERMDISP and

275 PERMANOVA programs of the Vegan package *Adonis* (Dixon 2003) with 1000 permutations.

276 Two models were used: the effects of the explanatory variables "Subspecies" and "Sex"

277 were tested on the whole dataset (counts ~ Subspecies + Sex + Subspecies\*Sex) and the

278 effects of the explanatory variables "Choosiness" (two levels: Choosy and Non-Choosy) and

279 "Sex" were tested on *M. m. musculus* samples only (counts ~ Choosiness + Sex +

280 Choosiness\*Sex). All explanatory variables were defined as fixed factors in our models.

281

282 Differential expression analyses were conducted using *edgeR* (version 2.13) (McCarthy *et al.*

283 2012). The analyses compared the two subspecies ('*domesticus* Choosy' versus '*musculus*

284 Choosy' samples) or the two groups of *M. m. musculus* mice showing distinct behavioural

285 profiles ('*musculus* Choosy' versus '*musculus* Non-Choosy'). Since males and females display

286 similar patterns of mate preference within each population (e.g., Smadja *et al.* 2004b, 2015;

287 Smadja & Ganem 2005), we considered four biological replicates per categorial group (two

288 sexes x two populations). Only genes expressed in all samples but one of the same

289 categorial group were included, leading to a different number of genes considered for each

290 comparison. Because simulation analyses have shown the tendency of *edgeR* to identify a

291 high number of false differentially expressed genes (Soneson & Delorenzi 2013), the  
292 significance level was reduced to 0.01. Additionally, a false discovery rate (FDR) method was  
293 used to correct the threshold of *P*-values for multiple testing (Benjamini & Yekutieli 2001).  
294 The outcome of this analysis was a list of up- and down-regulated differentially expressed  
295 (DE) genes.

296

### 297 **Enrichment analyses**

298 Functional annotation of the significantly differentially expressed genes was performed  
299 using the functional classification program *DAVID* (Version 6.8) (Huang *et al.* 2009).  
300 Enrichment analyses were performed for DE genes between the two subspecies and for DE  
301 genes between 'Choosy' and 'Non-Choosy' samples within *M. m. musculus*. Functional  
302 categories were considered significantly enriched when  $FDR < 0.05$  and we analysed  
303 functional annotation clusters, which group similar functional categories together, to  
304 identify enriched biological functions. In the case of VR genes, we ran an independent hyper-  
305 geometric test, as some of the genes are not annotated as VR genes in the database. In all  
306 cases, the reference background genome included only expressed genes for each  
307 comparison, instead of using the whole reference mouse genome.

308

### 309 **Phylogenetic and clustering analyses**

310 To map DE genes onto the phylogenies of vomeronasal receptor gene family 1 (V1R) and 2  
311 (V2R), we first reconstructed these gene phylogenies using available coding sequences of  
312 V1R and V2R (191 and 110 sequences respectively). Sequences were aligned using *MACSE*  
313 (Multiple Alignment of Coding Sequences; Ranwez *et al.* 2011), with manual adjustments.  
314 *jModelTest* v0.1.1 (Posada 2008) identified GTR+G+I (General Time Reversible + Gamma +  
315 Invariable sites) as the model of sequence evolution that best described the observed  
316 pattern of sequence variation for both V1R and V2R, and this model was used to perform

317 maximum likelihood phylogenetic analyses using *RAxML* version 8 (Stamatakis 2014). We  
318 then reported positions of DE genes identified previously between the two subspecies and  
319 between 'Choosy' and 'Non-Choosy' samples within *M. m. musculus* on the phylogenetic  
320 trees of the V1R and V2R families.

321

322 Since the phylogenetic analyses indicated a close relationship among five V2R genes that  
323 were found to be differentially expressed between the 'Choosy' and 'Non-Choosy' samples  
324 within *M. m. musculus*, we assessed the probability of phylogenetic clustering using a  
325 randomisation procedure. We randomly sampled five receptor genes in the phylogenetic  
326 tree of vomeronasal receptor family 2 and computed the total genetic distance among them  
327 using the *ape* R package (Paradis *et al.* 2004). We performed this random sampling 500,000  
328 times to obtain an expected distribution of genetic distance. The observed value of genetic  
329 distance among the five DE V2R genes was then compared to this distribution.

330

331 Since all VR genes belong to only two different gene families, they share a significant amount  
332 of genetic identity. Since mapping was based on sequence identity, we were concerned that  
333 reads mapping to one truly differentially expressed gene may also map to other VR paralogs  
334 with close sequence identity, leading to false positive results in the clustering analysis.

335 Therefore, for each VR gene found to be differentially expressed in the '*musculus* Choosy'  
336 versus '*musculus* Non-Choosy' comparison, we retrieved the 10 closest paralogs with the  
337 program *cd-hit-est* (<http://weizhong-lab.ucsd.edu/cd-hit/>). Sequences were aligned using  
338 the program *MACSE* (Ranwez *et al.* 2011) and positions that were specific to the  
339 differentially expressed genes were recorded. Then, using mapping information, we  
340 retrieved the coverage at each position for each of the DE genes in the relevant samples,  
341 and checked that positions specific to individual genes were covered at the same level as  
342 shared positions.

343

## 344 RESULTS

345

### 346 Variation in gene expression among samples

347 All samples were sequenced successfully, except one from the '*Musculus* Non-Choosy' group  
348 ('*Musculus* Non-Choosy population2 Females', Table S1) due to failure in library  
349 construction. For the eleven sequenced samples, the percentage of mapped reads per  
350 sample was between 79% and 85%, and the total number of transcripts per sample varied  
351 from 33457 to 39770 (Table S2). Table S3 reports the raw and normalised counts for all  
352 genes and all samples.

353

354 The MDS plots, representing variation in overall gene coverage among samples, indicated a  
355 clear separation along dimension 1 between samples of the two subspecies (Figure 1A) and  
356 some degree of separation between 'Choosy' and 'Non-Choosy' samples within *M. m.*  
357 *musculus* (Figure 1B). Multivariate analysis of variance showed that the overall gene  
358 expression levels differed significantly between *M. m. musculus* and *M. m. domesticus*  
359 samples but not between sexes (perMANOVA Adonis test:  $F_{\text{subspecies}} = 6.423$ ;  $P = 0.004$ ;  $F_{\text{sex}} =$   
360  $0.804$ ;  $P = 0.547$ ;  $F_{\text{subspecies} \times \text{sex}} = 0.322$ ;  $P = 0.916$ ). The trend for a separation between  
361 'Choosy' and 'Non-Choosy' samples within *M. m. musculus* was not statistically significant  
362 (perMANOVA Adonis test:  $F_{\text{choosiness}} = 2.424$ ;  $P = 0.089$ ;  $F_{\text{sex}} = 0.782$ ;  $P = 0.537$ ;  $F_{\text{choosiness} \times \text{sex}} =$   
363  $0.506$ ;  $P = 0.835$ ).

364

365 The number of expressed VR genes varied slightly among samples (Table S4a) but remained  
366 comparable (range between 197 to 252), with no significant difference between samples  
367 from the most divergent groups, i.e. the two subspecies (Wilcoxon rank sum test,  $P = 0.886$ ).

368 We also detected expression of Olfactory Receptors (OR) in the vomeronasal organ samples



369 (Tables S3 and S4). The mean levels of expression of VR and OR genes were lower than the  
370 mean coverage for the rest of the genes (mean CPM<sub>VR</sub> = 3.414; mean CPM<sub>OR</sub> = 0.804; mean  
371 CPM<sub>other\_genes</sub> = 27.198), with ORs particularly lowly expressed (Table S4b and Figure S2). In  
372 contrast, the mean of median CPM values across all samples was higher for VRs than for the  
373 other genes (mean median<sub>VR</sub> = 0.57; mean median<sub>other\_genes</sub> = 0.29) (Figure S2).

374

375 Finally, we report expression of some mouse lipocalins in our VNO samples (Table S3). Seven  
376 Odorant Binding Proteins (OBP) were expressed in all samples (including *Obp3-p* annotated  
377 as a pseudogene in the mouse reference genome Stopková *et al.* 2016). Among the 12 Major  
378 Urinary Proteins (MUP) annotated as functional genes and pseudogenes in the reference  
379 genome and that belong to Group A (i.e. the ancestral group of MUPs for which mapping  
380 information is reliable, Logan *et al.* 2008), we detected expression of three MUPs in all  
381 samples (*Mup4*, *Mup5* and to a lesser extent *Mup6*), and four other MUPs were lowly  
382 expressed, mostly in *M. m. musculus* samples (*Mup3*, *Mup20* (also known as Darcin), *Mup-*  
383 *ps19* and *Mup-ps20*).

384

### 385 **Differential expression among samples**

386 Differential expression analyses were performed on the comparison between the two  
387 subspecies and the comparison between 'Choosy' and 'Non-Choosy' samples within *M. m.*  
388 *musculus*. DE genes showed log fold changes from 1 to 10 and average coverage between 1  
389 and 12 logCPM (counts per million) (Figure 2). We found a large number of genes (2624)  
390 differentially expressed between *M. m. domesticus* and *M. m. musculus* samples (Figure 2A,  
391 Table S5a). In contrast, we only found 236 genes differentially expressed between the  
392 'Choosy' and 'Non-Choosy' *M. m. musculus* samples (Figure 2B, Table S5b). For this  
393 comparison, there were more differentially expressed genes up-regulated in the 'Choosy'  
394 samples (156 genes) than the 'Non-Choosy' samples (90 genes) (Table S5b), and in general

395 the values of log fold change were also higher in the 'Choosy' samples (Figure 2B).  
396  
397 We identified 90 VR genes and one OR gene (*Olf460*) differentially expressed between the  
398 two subspecies (Table S5a). We also identified three lipocalins that were differentially  
399 expressed between the two taxa (Table S5a): two MUP genes from Group A (*Mup6* and  
400 *Mup20*) expressed mostly in *M. m. musculus* samples, and one OBP gene (*Obp6*) only  
401 expressed in *M. m. domesticus* samples. More generally, functional analyses revealed an  
402 enrichment in 35 functional categories (GO terms) in the set of genes differentially  
403 expressed between the two subspecies, with the most enriched clusters of annotation terms  
404 corresponding to signal-related functions, vomeronasal receptors of type 2 (V2Rs) and other  
405 G-protein-coupled receptors (Table S6a). For the comparison between 'Choosy' and 'Non-  
406 Choosy' samples within *M. m. musculus*, we identified 34 enriched functional categories,  
407 with the most enriched clusters of annotation terms corresponding to immune-related  
408 functions, signal-related functions and sensory perception of smell (Table S6b). Among the  
409 genes found to be differentially expressed between 'Choosy' and 'Non-Choosy' samples, we  
410 found five V2R genes (*V2r17*, *V2r69*, *V2r70*, *V2r73* and *V2r75*), which corresponds to a  
411 significant enrichment of this category of genes (hypergeometric test:  $P < 0.01$  with 5 out of  
412 140 in a sample of 236 DE genes). Among the five DE V2Rs, four were up-regulated in the  
413 'Choosy' samples and one in the 'Non-Choosy' samples (Figure 3). Differentially expressed  
414 genes between 'Choosy' and 'Non-Choosy' samples within *M. m. musculus* also included two  
415 OR genes (*Olf15* and *Olf449*) but no members from family 1 of the VR genes (V1R) (Table  
416 S5b).

417

#### 418 **Phylogenetic relationships among differentially expressed vomeronasal receptor genes**

419 When plotting the differentially expressed genes on the vomeronasal receptor phylogenies,  
420 we observed that genes differentially expressed between the two subspecies mapped

421 almost everywhere on the V1R and V2R phylogenies (Figure 4A and 4B). In contrast, genes  
422 differentially expressed between 'Choosy' and 'Non-Choosy' samples only mapped to family  
423 2 of vomeronasal receptors (Figure 4B). One DE VR gene, *V2r17*, belongs to clade 6 of the  
424 V2R family. Interestingly, the other four DE VR genes belong to the same phylogenetic clade,  
425 clade 3 (genes *V2r69*, *V2r70*, *V2r73* and *V2r75*). The phylogenetic distance between the five  
426 DE VR genes was found to be significantly smaller than expected by chance (random  
427 sampling test, observed phylogenetic distance = 3.69;  $p < 0.004$  and Figure S3).

428

429 The fact that closely related VR genes were found to be differentially expressed between the  
430 'Choosy' and 'Non-Choosy' samples within *M. m. musculus* could reflect a bias in mapping  
431 due to sequence similarity among closely-related paralogs. Indeed, these four VR belong to  
432 the same genomic cluster on chromosome 7, and are probably tandem-repeat paralogs with  
433 relatively high sequence similarity. However, we found no evidence for such a bias. Although  
434 the coverage along the gene was variable, there was no difference between the coverage of  
435 conserved and variable regions in any gene, and there were numerous variable sites with  
436 high coverage (Figure S4). This was true even when DE genes were compared to their most  
437 closely-related paralogs (same phylogenetic cluster). For example, analysis of coverage  
438 information along *V2r70* (differentially expressed) and its closest relative *V2r74*  
439 (ENSMUSG00000090774) (not differentially expressed) showed prominent differences in  
440 coverage between these two genes for both variable and conserved regions, which indicates  
441 that the mapping was able to differentiate reads from different but closely related VR genes.

442

443 **DISCUSSION**

444 Changes in expression are expected to play an important role in the evolution and  
445 divergence of traits involved in reproductive isolation. Here we have used RNA-seq on whole  
446 vomeronasal organs collected from wild mice sampled in Denmark to test for the role of  
447 changes in expression of vomeronasal receptors in the behavioural shift toward olfactory-  
448 based assortative mating observed in the hybrid zone between the two European subspecies  
449 of the house mouse. Overall expression levels did not vary significantly between males and  
450 females in our dataset, a result in agreement with previous comparisons of male and female  
451 mouse VNO transcriptomes (Ibarra-Soria *et al.* 2014). Since assortative mate preferences are  
452 displayed by both males and females in the studied mouse populations (Smadja *et al.* 2004b;  
453 Smadja & Ganem 2005), we did not analyse differential expression between males and  
454 females further. Overall expression levels differed between the two subspecies, and we  
455 found a very large number of genes differentially expressed between the two taxa (more  
456 than 2,000 in total, and 90 VRs), consistent with the degree of divergence between the two  
457 taxa (Phifer-Rixey *et al.* 2014). In contrast, only 236 genes were found repeatedly  
458 differentially expressed between the *M. m. musculus* populations showing contrasting  
459 patterns of mate preference. Among these DE genes between 'Choosy' and 'Non-Choosy'  
460 samples, we found an overrepresentation of genes belonging to family 2 of vomeronasal  
461 receptors, known to play a key role in pheromone recognition in mice (Chamero *et al.* 2012).  
462 Interestingly, four out of five of these differentially expressed VRs belong to the same  
463 phylogenetic cluster. In the following paragraphs, we focus on discussing these findings  
464 related to expression divergence between populations of *M. m. musculus* showing  
465 contrasting mate preference behaviours.

466

467

468 *How confident can we be that differences in expression are associated with differences in*  
469 *behaviour?*

470 Comparative transcriptomics has previously been used to reveal candidate genes for  
471 behavioural differences or to reconstruct preference-gene networks (reviewed in Wilkinson  
472 *et al.* 2015). In our study, how confident can we be that these differences in gene expression  
473 are, at least in part, associated with the observed behavioural shift toward assortative  
474 mating in the populations nearer to the hybrid zone? First, the use of several biological  
475 replicates for each of the two types of natural populations known to differ in their mate  
476 preference patterns minimizes the risk of observing population-specific differences that  
477 could reflect different regimes of selection or demographic events among populations.  
478 Second, we sampled *M. m. musculus* populations with contrasting mate preference  
479 behaviours but are geographically close in Denmark. The different populations thus share a  
480 very recent common ancestor, have similar low rates of introgression from *M. m.*  
481 *domesticus* and live in similar environments. Our aim was to minimize any other possible  
482 natural sources of divergence between these populations and therefore maximize the  
483 chance that genetic/expression differences reflect, at least in part, the known behavioural  
484 difference between the groups. Since VR expression at the adult stage is known to be  
485 influenced by postnatal (from birth to 19 days old) exposure to odours (environment,  
486 parental imprinting) at least at the kin recognition level (Broad & Keverne 2012), we decided  
487 to study wild-caught mice to ensure assessment of patterns of gene expression as close as  
488 possible to natural populations, i.e. potentially influenced by a combination of innate and  
489 imprinted mechanisms. However, to homogenise replicates further and avoid any possible  
490 variation in gene expression due to differing physiological conditions, we maintained wild-  
491 caught mice in the laboratory before dissection to control for the main factors known to  
492 influence expression in general and VR expression in particular (diet, temperature,  
493 photoperiod, the olfactory environment, age range). Although we could not strictly control

494 for the age of these wild-caught individuals, we selected only mice trapped as young adults  
495 expected to fall in the age range corresponding to the peaked plateau of VR expression in  
496 adult mice (Zhang *et al.* 2010). This ensured that age-related variation in VR expression  
497 would not strongly influence our results.

498 Considering that the experiment was performed using wild animals from natural  
499 populations, we cannot completely exclude other sources of variation in the observed  
500 profiles of gene expression. However, we consider that this series of experimental settings,  
501 combined with a targeted approach to candidate genes likely to be involved in the behaviour  
502 of interest, offers optimal conditions to detect association between assortative mate  
503 preferences and differential gene expression.

504

#### 505 *Identity and functions of differentially expressed genes*

506 Our main objective was to assess patterns of expression of vomeronasal receptors.  
507 However, we identified several other classes of genes showing repeated patterns of  
508 differential expression between 'Choosy' and 'Non-Choosy' samples. Several of these genes  
509 have physiological functions related to signal transportation, perception and transduction.  
510 Enrichment analyses showed an overrepresentation of genes involved in regulation of  
511 odorant selectivity in the extracellular space and olfactory lumen. This includes genes coding  
512 for androgen-binding proteins (ABP), previously suggested to act as subspecific signals in  
513 mice (Laukaitis *et al.* 1997; Bimova *et al.* 2011), and some organic anion transporters (OAT)  
514 expressed in the mouse olfactory mucosa. The latter are potentially involved in odour  
515 signalling by binding and transporting organic anions present in urine (Kaler *et al.* 2006).  
516 Other enriched categories include genes involved in regulation of signal transduction (cyclic-  
517 nucleotide-gated channel (CNG) gene family; Francia, Pifferi, Menini, & Tirindelli, 2014). We  
518 also identified two olfactory receptors differentially expressed between the two behavioural  
519 groups. Although ORs are primarily expressed in the MOE, our results are congruent with a

520 recent study reporting expression of ORs in mouse VNOs (Ibarra-Soria *et al.* 2014). We did  
521 not detect differential expression of MUPs (from Group A) or OBPs between 'Choosy' and  
522 'Non-Choosy' samples, but could confirm previous findings evidencing their expression in  
523 VNO tissues (Utsumi *et al.* 1999; Ibarra-Soria *et al.* 2014; Stopková *et al.* 2016) and  
524 expression divergence at some MUP and OBP genes between the two subspecies (although  
525 not affecting the same genes as in Stopková *et al.* 2016). These lipocalins expressed in the  
526 VNO may play a role in sequestering volatile pheromones and possibly transporting them to  
527 vomeronasal receptors (Sharrow *et al.* 2002; Stopková *et al.* 2016). Interestingly, the  
528 diversity of genes and functions identified as differentially expressed between 'Choosy' and  
529 'Non-Choosy' samples suggests that the evolution of assortative mating involves a complex  
530 network of VNO genes that will need further investigation. Additionally, genes expressed in  
531 other olfactory tissues (MOE) or involved in the integration of the signal in the brain, which  
532 we did not study here, may also play an important role in the expression of assortative mate  
533 preference.

534

535 As far as VR are concerned, we identified five receptor genes (*V2r17*, *V2r69*, *V2r70*, *V2r73*  
536 and *V2r75*) that showed repeated differential expression between the two behavioural  
537 groups. Since in most cases only one functional receptor is expressed per neuron in the  
538 vomeronasal organ (Dalton & Lomvardas 2015; Ishii & Mombaerts 2011), the observed  
539 expression divergence at individual VRs could result from differences in transcript level per  
540 neuron (Young *et al.* 2003), but also from differences in the numbers of neurons expressing  
541 a given receptor, a parameter that shapes the olfactory map in the accessory olfactory bulb  
542 and provides to the brain the signal for a behavioural response (Mombaerts 2004; Bear *et al.*  
543 2016). Since we used RNA-seq on whole vomeronasal organs, our experiment does not allow  
544 us to distinguish these two levels of expression, but is informative about the overall pattern  
545 of variation and divergence at individual VRs that likely influence odour perception.

546 Four of the differentially expressed VR were found to be up-regulated in the 'Choosy'  
547 samples and one was up-regulated in the 'Non-Choosy' samples. A majority of VR genes  
548 upregulated in the 'Choosy' samples is consistent with the idea that some receptors are  
549 recruited in the zone of contact with the other subspecies to allow the expression of  
550 assortative mate preferences.

551 In contrast to differentially expressed VRs between the two subspecies belonging to both  
552 families of VRs (family 1 and family 2), expression differentiation within *M. m. musculus*  
553 between 'Choosy' and 'Non-Choosy' samples only affects members of family 2 (V2R). This  
554 corresponds to a significant enrichment of this category of genes among all DE genes.

555 Traditionally, V2Rs were thought to be involved primarily in pheromone recognition (as  
556 opposed to V1Rs that were thought to be involved in the detection of the physiological  
557 status of other animals). Recent findings have shown that individual V2Rs are specifically  
558 tuned to detect the nature of the signal emitter (e.g. conspecifics, heterospecifics, individual  
559 recognition), whereas individual V1Rs are activated by cues from multiple species (Isogai *et al.*  
560 *al.* 2011). Additionally, V2R-positive Vomeronasal Sensory Neurons detect members of  
561 several large peptide or protein families (in particular MHC class I peptides and some Major  
562 Urinary Proteins (MUPs) secreted in urine) that are thought to inform the receiver about the  
563 genetic identity of a signaler (Hurst *et al.* 2001; Leinders-Zufall *et al.* 2004; Papes *et al.* 2010;  
564 Chamero *et al.* 2012). In agreement with this, urines have been shown to diverge in their  
565 composition of MUPs between the two subspecies (Hurst *et al.* 2017; Stopková *et al.* 2007)  
566 and between 'Choosy' versus 'Non-Choosy' samples within *M. m. musculus* (Hurst *et al.*  
567 2017), which establishes MUPs as potential candidates acting as subspecific signals and  
568 potentially recognised by V2Rs. Since our study also confirms previous findings that indicate  
569 that some MUPs are also expressed in the VNO, a future challenge will be to understand  
570 better the respective role of VNO-expressed MUPs and MUPs expressed in the liver and  
571 excreted in urine in mouse communication.



572

573 *Phylogenetic clustering and functional specialisation of differentially expressed vomeronasal*  
574 *receptors*

575 We found that four out of five of the differentially expressed V2Rs belong to the same  
576 phylogenetic cluster (*V2r69*, *V2r70*, *V2r73* and *V2r75* in clade 3 of the V2R family). Although  
577 these four DE genes are in the same genomic cluster on chromosome 7 and share a  
578 relatively high degree of sequence similarity, we have evidence to exclude the possibility of a  
579 methodological artefact. We did not find any evidence for a bias when aligning these DE  
580 genes with other closely-related paralogs and comparing coverage information in specific  
581 versus shared positions along these genes. Moreover, some of the closely-related paralogs  
582 belonging to the same genomic cluster as the DE genes did not show significant differential  
583 expression. If this pattern of phylogenetic clustering of DE VRs is true, it suggests that the  
584 evolution of assortative mate preference in *M. m. musculus* could involve the specialisation  
585 of a group of closely-related receptors in the clade 3 of V2Rs. This hypothesis is consistent  
586 with the study by Isogai *et al.* (2011) which showed that distinct mouse VR receptor  
587 subfamilies have evolved towards the specific recognition of certain animal groups  
588 (heterospecific signals from predators or sympatric closely-related species/subspecies, male  
589 or female conspecific signals) or chemical structures (volatile steroids, non-volatile proteins).  
590 Interestingly, the VNOs analysed in their study were from *M. m. domesticus* (laboratory  
591 mice), and when tested using *M. m. musculus* scents, several V2Rs were activated including  
592 *V2r17* (DE gene in our study) and its close relatives and *V2r66* (belonging to the same clade 3  
593 as the four VRs differentially expressed in our study). Other receptors identified as  
594 differentially expressed in our study could not be individually functionally characterised.  
595 Based on these results, these authors and others (Chamero *et al.* 2012; Wynn *et al.* 2012)  
596 hypothesized a role of these VRs in reproductive isolation. An interesting follow-up from our  
597 work would be to apply the methods developed by Isogai *et al.* to test whether the same set

598 of closely-related V2Rs, identified by these authors as activated by *M. m. musculus* scents  
599 when expressed in *M. m. domesticus* VNOs, would be activated by *M. m. domesticus* stimuli  
600 when expressed by *M. m. musculus* VNOs, as our RNA-seq results suggest.

601

#### 602 *Evolution of expression changes and the role of selection*

603 Regulatory changes in genes involved in reproduction are thought to be prime targets for  
604 divergence during speciation. They are expected to evolve under natural and/or sexual  
605 selection. Previous studies had shown that expression divergence in testis-associated genes  
606 seems to establish during a later phase of the speciation process when reproductive  
607 isolation is complete (Voolstra, Tautz, Farbrother, Eichinger, & Harr, 2007; Bryk, Somel,  
608 Lorenc, & Teschke, 2013). In our study, we provide evidence for expression divergence in  
609 VNO-related genes between populations within the *M. m. musculus* subspecies, in a context  
610 of secondary contact with the subspecies *M. m. domesticus*. This result may suggest that  
611 expression divergence at these genes has evolved rapidly in the ‘Choosy’ populations from  
612 the border of the hybrid zone. Whatever the source of expression divergence (new  
613 mutations, standing variation, hybridisation), changes in expression related to the evolution  
614 of assortative mate preferences are expected, in the context of reinforcement, to evolve  
615 rapidly under diverging selection (natural and sexual selection against hybridisation in the  
616 house mouse system, Britton-Davidian *et al.* 2005; Latour *et al.* 2014). Although  
617 accumulating evidence suggests a predominantly neutral model of gene expression  
618 evolution (e.g. in house mice, Staubach, Teschke, Voolstra, Wolf, & Tautz, 2010), positive  
619 selection has also been shown to shape changes in expression for some genes and lineages  
620 (e.g. Wittkopp, Haerum, & Clark, 2008; Jeukens & Bernatchez, 2012; Chapman *et al.* 2013).  
621 In a recent study, we identified several genomic regions surrounding VR genes in the house  
622 mouse with signatures of positive selection specifically in populations from the border of the  
623 hybrid zone that display assortative mate preference (same samples as the ‘Choosy’ *M. m.*

624 *musculus* samples analysed in the present study). Interestingly, one of these genomic  
625 regions corresponds to the downstream and putative regulatory region of the cluster  
626 containing the candidate VRs identified in the present study, in the vicinity of *V2r75* (Smadja  
627 *et al.* 2015). Another study on wild-derived strains of the two subspecies has suggested that  
628 V2Rs identified by Isogai *et al.* (2011) as detecting subspecific cues are prone to evolve  
629 under positive selection, in contrast to V2Rs detecting other classes of signals (Wynn *et al.*  
630 2012). Analysing genome-wide patterns of nucleotide polymorphism and divergence in the  
631 same biological samples as the ones used in the present study should allow us to confirm  
632 whether some differentially expressed vomeronasal receptors or their *cis*-regulatory regions  
633 evolve under divergent selection in the populations displaying assortative mate preferences.  
634  
635 To conclude, this study suggests that changes in expression at key receptors for pheromone  
636 recognition may play an important role in the evolution of assortative mating and reinforced  
637 sexual isolation in the house mouse. Evidence for differential expression at some closely-  
638 related vomeronasal receptors between populations within *M. m. musculus* that do or do  
639 not display assortative mate preferences (close to or more distant from the hybrid zone,  
640 respectively) implicates them in this behavioural shift. This is consistent with previous  
641 findings suggesting the specialisation of this group of receptors in the recognition of odorant  
642 signals that may allow subspecies recognition and assortative mating.

643

644 ACKNOWLEDGEMENTS

645 We are very grateful to Ingrid Jakob, former engineer at Centre des Sciences du Goût et de  
646 l'Alimentation (UMR 1324), University of Burgundy, Dijon, France, for her advice on  
647 dissecting mouse vomeronasal organs. We would also like to thank very much Josette  
648 Catalan, Marco Perriat-Sanguinet, Yamin Latour for their help in collecting mice in Denmark  
649 and the Danish farmers for their hospitality, as well as Giveskud zoo personnel. We thank  
650 Janice Britton-Davidian, Josette Catalan and Marco Perriat-Sanguinet for their help in  
651 dissecting mice. We thank the platform "High throughput qPCR" of the University of  
652 Montpellier for having given us access to an Agilent instrument to assess RNA sample  
653 quality. Finally, we thank Roger Butlin for his comments on a previous version of this  
654 manuscript. The research leading to these results has received funding from the European  
655 Union's Seventh Framework Programme [FP7/2007-2013] – Marie Curie European  
656 Reintegration Grant (ERG), under Grant agreement n°PERG06-GA-2009-251008, as well as  
657 from the Agence Nationale pour la Recherche (ANR) under Grant agreement No.  
658 2010BLAN171401-AssortMate. This is publication ISEM 2017-114.

659

660

661 REFERENCES

- 662 Albrechtova J, Albrecht T, Baird SJE *et al.* (2012) Sperm-related phenotypes implicated in  
663 both maintenance and breakdown of a natural species barrier in the house mouse.  
664 *Proceedings of the Royal Society B-Biological Sciences*, **279**, 4803–4810.
- 665 Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance.  
666 *Austral Ecology*, **26**, 32–46.
- 667 Bailey RI, Innocenti P, Morrow EH, Friberg U, Qvarnström A (2011) Female *Drosophila*  
668 *melanogaster* gene expression and mate choice: the X chromosome harbours  
669 candidate genes underlying sexual isolation. *PloS one*, **6**, e17358.
- 670 Bear DM, Lassance J-M, Hoekstra HE, Datta SR (2016) The Evolving Neural and Genetic  
671 Architecture of Vertebrate Olfaction. *Current Biology*, **26**, R1039–R1049.
- 672 Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing  
673 under dependency. *Annals of Statistics*, **29**, 1165–1188.
- 674 Bhattacharyya T, Gregorova S, Mihola O *et al.* (2013) Mechanistic basis of infertility of mouse  
675 intersubspecific hybrids. *Proceedings of the National Academy of Sciences*, **110**, E468–  
676 E477.
- 677 Bimova B, Macholan M, Baird SJE *et al.* (2011) Reinforcement selection acting on the  
678 European house mouse hybrid zone. *Molecular Ecology*, **20**, 2403–2424.
- 679 Boursot P, Din W, Anand R *et al.* (1996) Origin and radiation of the house mouse:  
680 mitochondrial DNA phylogeny. *J. Evol. Biol.*, **9**, 391–415.
- 681 Britton-Davidian J, Fel-Clair F, Lopez J, Alibert P, Boursot P (2005) Postzygotic isolation  
682 between the two European subspecies of the house mouse: estimates from fertility  
683 patterns in wild and laboratory-bred hybrids. *Biol. J. Linn. Soc.*, **84**, 379–393.
- 684 Broad K, Keverne E (2012) The Post-Natal Chemosensory Environment Induces Epigenetic  
685 Changes in Vomeronasal Receptor Gene Expression and a Bias in Olfactory Preference.

686 *Behavior Genetics*, **42**, 461–471.

687 Bryk J, Somel M, Lorenc A, Teschke M (2013) Early gene expression divergence between  
688 allopatric populations of the house mouse (*Mus musculus domesticus*). *Ecology and*  
689 *evolution*, **3**, 558–68.

690 Cadiou H, Aoudé I, Tazir B *et al.* (2014) Postnatal odorant exposure induces peripheral  
691 olfactory plasticity at the cellular level. *The Journal of neuroscience : the official journal*  
692 *of the Society for Neuroscience*, **34**, 4857–70.

693 Carroll SB, Grenier JK, Weatherbee SD (2005) *From DNA to diversity : molecular genetics and*  
694 *the evolution of animal design*. Blackwell Pub.

695 Cassone BJ, Mouline K, Hahn MW *et al.* (2008) Differential gene expression in incipient  
696 species of *Anopheles gambiae*. *Molecular ecology*, **17**, 2491–504.

697 Chamero P, Leinders-Zufall T, Zufall F (2012) From genes to social communication: molecular  
698 sensing by the vomeronasal organ. *Trends in Neurosciences*, **35**, 597–606.

699 Chan YF, Marks ME, Jones FC *et al.* (2010a) Adaptive Evolution of Pelvic Reduction in  
700 Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science*, **327**, 302–305.

701 Chan YF, Marks ME, Jones FC *et al.* (2010b) Adaptive Evolution of Pelvic Reduction in  
702 Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science*, **327**, 302–305.

703 Chapman MA, Hiscock SJ, Filatov DA (2013) Genomic Divergence during Speciation Driven by  
704 Adaptation to Altitude. *Molecular Biology and Evolution*.

705 Chung H, Loehlin DW, Dufour HD *et al.* (2014) A Single Gene Affects Both Ecological  
706 Divergence and Mate Choice in *Drosophila*. *Science*, **343**, 1148–1151.

707 Cucchi T, Auffray J-C, Vigne J-D (2012) History of house mouse synanthropy and dispersal in  
708 the near east and europe: a zooarchaeological insight. In: *Evolution of the house mouse*  
709 (eds Macholan M, Baird SJE, P M, Pialek J), pp. 65–93. Cambridge University Press.

710 Dalton RP, Lomvardas S (2015) Chemosensory receptor specificity and regulation. *Annual*

711 *review of neuroscience*, **38**, 331–49.

712 Davidson JH, Balakrishnan CN (2016) Gene Regulatory Evolution During Speciation in a  
713 Songbird. *G3&#58; Genes/Genomes/Genetics*, **6**, 1357–1364.

714 Dixon P (2003) VEGAN, a package of R functions for community ecology. *Journal of*  
715 *Vegetation Science*, **14**, 927–930.

716 Etges WJ (2014) No boundaries: genomes, organisms, and ecological interactions  
717 responsible for divergence and reproductive isolation. *The Journal of heredity*, **105**  
718 **Suppl**, 756–70.

719 Eyres I, Jaquiéry J, Sugio A *et al.* (2016) Differential gene expression according to race and  
720 host plant in the pea aphid. *Molecular Ecology*, **25**, 4197–4215.

721 Fortes-Marco L, Lanuza E, Martinez-Garcia F (2013) Of Pheromones and Kairomones: What  
722 Receptors Mediate Innate Emotional Responses? *The Anatomical Record*, **296**, 1346–  
723 1363.

724 Francia S, Pifferi S, Menini A, Tirindelli R (2014) *Vomeronasal Receptors and Signal*  
725 *Transduction in the Vomeronasal Organ of Mammals*. CRC Press/Taylor & Francis.

726 Fujii T, Fujii T, Namiki S *et al.* (2011) Sex-linked transcription factor involved in a shift of sex-  
727 pheromone preference in the silkworm *Bombyx mori*. *Proceedings of the National*  
728 *Academy of Sciences of the United States of America*, **108**, 18038–43.

729 Ganem G, Litel C, Lenormand T (2008) Variation in mate preference across a house mouse  
730 hybrid zone. *Heredity*, **100**, 594–601.

731 Haerty W, Singh RS (2006) Gene regulation divergence is a major contributor to the  
732 evolution of Dobzhansky-Muller incompatibilities between species of *Drosophila*.  
733 *Molecular biology and evolution*, **23**, 1707–14.

734 Hansen KD, Irizarry RA, Wu Z (2012) Removing technical variability in RNA-seq data using  
735 conditional quantile normalization. *Biostatistics (Oxford, England)*, **13**, 204–16.

736 Hoang K, Matzkin LM, Bono JM (2015) Transcriptional variation associated with cactus host  
737 plant adaptation in *Drosophila mettleri* populations. *Molecular Ecology*, n/a-n/a.

738 Hoekstra HE, Coyne JA (2007) The locus of evolution: Evo devo and the genetics of  
739 adaptation. *Evolution*, **61**, 995–1016.

740 Horth L (2007) Sensory genes and mate choice: Evidence that duplications, mutations, and  
741 adaptive evolution alter variation in mating cue genes and their receptors. *Genomics*,  
742 **90**, 159–175.

743 Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large  
744 gene lists using DAVID bioinformatics resources. *Nature protocols*, **4**, 44–57.

745 Huestis DL, Marshall JL (2009) From Gene Expression to Phenotype in Insects: Non-  
746 microarray Approaches for Transcriptome Analysis. *BioScience*, **59**, 373–384.

747 Hurst JL, Beynon RJ, Armstrong SD *et al.* (2017) Molecular heterogeneity in major urinary  
748 proteins of *Mus musculus* subspecies: potential candidates involved in speciation.  
749 *Scientific Reports*, **7**, 44992.

750 Hurst JL, Payne CE, Nevison CM *et al.* (2001) Individual recognition in mice mediated by  
751 major urinary proteins. *Nature*, **414**, 631–634.

752 Ibarra-Soria X, Levitin M, Logan D (2013) The genomic basis of vomeronasal-mediated  
753 behaviour. *Mammalian Genome*, 1–12.

754 Ibarra-Soria X, Levitin MO, Saraiva LR, Logan DW (2014) The Olfactory Transcriptomes of  
755 Mice. (J Zhang, Ed,). *PLoS genetics*, **10**, e1004593.

756 Isogai Y, Si S, Pont-Lezica L *et al.* (2011) Molecular organization of vomeronasal  
757 chemoreception. *Nature*, **478**, 241–245.

758 Kaler G, Truong DM, Sweeney DE *et al.* (2006) Olfactory mucosa-expressed organic anion  
759 transporter, Oat6, manifests high affinity interactions with odorant organic anions.  
760 *Biochemical and biophysical research communications*, **351**, 872–6.



761 Karn RC, Young JM, Laukaitis CM (2010) A Candidate Subspecies Discrimination System  
762 Involving a Vomeronasal Receptor Gene with Different Alleles Fixed in *M. m.*  
763 *domesticus* and *M. m. musculus*. *PLoS ONE*, **5**, 11.

764 Keller M, Baum MJ, Brock O, Brennan PA, Bakker J (2009) The main and the accessory  
765 olfactory systems interact in the control of mate recognition and sexual behavior.  
766 *Behavioural Brain Research*, **200**, 268–276.

767 King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. *Science*  
768 (*New York, N.Y.*), **188**, 107–16.

769 Latour Y, Perriat-Sanguinet M, Caminade P *et al.* (2014) Sexual selection against natural  
770 hybrids may contribute to reinforcement in a house mouse hybrid zone. *Proceedings of*  
771 *the Royal Society B: Biological Sciences*, **281**.

772 Laukaitis CM, Crister ES, Karn RC (1997) Salivary androgen-binding protein (ABP) mediates  
773 sexual isolation in *Mus musculus*. *Evolution*, **51**, 2000–2005.

774 Leinders-Zufall T, Brennan P, Widmayer P *et al.* (2004) MHC class I peptides as chemosensory  
775 signals in the vomeronasal organ. *Science*, **306**, 1033–1037.

776 Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or  
777 without a reference genome. *BMC bioinformatics*, **12**, 323.

778 Liberles SD (2014) Mammalian Pheromones. *Annual Review of Physiology*, **76**, 151–175.

779 Logan DW, Marton TF, Stowers L (2008) Species Specificity in Major Urinary Proteins by  
780 Parallel Evolution. *PLoS ONE*, **3**, e3280.

781 Mack KL, Campbell P, Nachman MW (2016) Gene regulation and speciation in house mice.  
782 *Genome Research*, **26**, 451–461.

783 Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing  
784 reads. *EMBnet.journal*, **17**, 10.

785 McBride CS, Baier F, Omondi AB *et al.* (2014) Evolution of mosquito preference for humans

786 linked to an odorant receptor. *Nature*, **515**, 222–227.

787 Mombaerts P (2004) Genes and ligands for odorant, vomeronasal and taste receptors.  
788 *Nature reviews. Neuroscience*, **5**, 263–78.

789 Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying  
790 mammalian transcriptomes by RNA-Seq. *Nat Meth*, **5**, 621–628.

791 Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences.  
792 *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 332–342.

793 Nosil P, Schluter D (2011) The genes underlying the process of speciation. *Trends in Ecology*  
794 *& Evolution*, **26**, 160–167.

795 Orr HA, Masly JP, Presgraves DC (2004) Speciation genes. *Current Opinion in Genetics &*  
796 *Development*, **14**, 675–679.

797 Ortiz-Barrientos D, Counterman BA, Noor MAF (2007) Gene expression divergence and the  
798 origin of hybrid dysfunctions. *Genetica*, **129**, 71–81.

799 Papes F, Logan DW, Stowers L (2010) The Vomeronasal Organ Mediates Interspecies  
800 Defensive Behaviors through Detection of Protein Pheromone Homologs. *Cell*, **141**,  
801 692–703.

802 Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R  
803 language. *Bioinformatics*, **20**, 289–290.

804 Pavey SA, Collin H, Nosil P, Rogers SM (2010) The role of gene expression in ecological  
805 speciation. In: *Year in Evolutionary Biology*, pp. 110–129. Blackwell Publishing, Oxford.

806 Pease JB, Guerrero RF, Sherman NA, Hahn MW, Moyle LC (2016) Molecular mechanisms of  
807 postmating prezygotic reproductive isolation uncovered by transcriptome analysis.  
808 *Molecular Ecology*, **25**, 2592–2608.

809 Phifer-Rixey M, Bomhoff M, Nachman MW (2014) Genome-Wide Patterns of Differentiation  
810 Among House Mouse Subspecies. *Genetics*, **198**, 283–297.

811 Ragland GJ, Almskaar K, Vertacnik KL *et al.* (2015) Differences in performance and  
812 transcriptome-wide gene expression associated with *Rhagoletis* (Diptera: Tephritidae)  
813 larvae feeding in alternate host fruit environments. *Molecular ecology*, **24**, 2759–76.

814 Ranwez V, Harispe S, Delsuc F, Douzery EJP (2011) MACSE: Multiple Alignment of Coding  
815 SEquences Accounting for Frameshifts and Stop Codons (WJ Murphy, Ed.). *PLoS ONE*, **6**,  
816 e22594.

817 Raufaste N, Orth A, Belkhir K *et al.* (2005) Inference of selection and migration in the danish  
818 house mouse hybrid zone. *Biological Journal of the Linnean Society*, **84**, 593–616.

819 Robinson MD, Oshlack A, Wang E *et al.* (2010) A scaling normalization method for  
820 differential expression analysis of RNA-seq data. *Genome Biology*, **11**, R25.

821 Seehausen O, Butlin RK, Keller I *et al.* (2014) Genomics and the origin of species. *Nat Rev*  
822 *Genet*, **15**, 176–192.

823 Sharrow SD, Vaughn JL, Zidek L, Novotny M V, Stone MJ (2002) Pheromone binding by  
824 polymorphic mouse major urinary proteins. *Protein Science*, **11**, 2247–2256.

825 Smadja C, Catalan J, Ganem G (2004a) Strong premating divergence in a unimodal hybrid  
826 zone between two subspecies of the house mouse. *Journal of Evolutionary Biology*, **17**,  
827 165–176.

828 Smadja C, Catalan J, Ganem G (2004b) Strong premating divergence in a unimodal hybrid  
829 zone between two subspecies of the house mouse. *Journal of Evolutionary Biology*, **17**,  
830 165–176.

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

833 Smadja C, Ganem G (2005) Asymmetrical reproductive character displacement in the house  
834 mouse. *Journal of Evolutionary Biology*, **18**, 1485–1493.

835 Smadja CM, Loire E, Caminade P *et al.* (2015) Seeking signatures of reinforcement at the

836 genetic level: a hitchhiking mapping and candidate gene approach in the house mouse.  
837 *Molecular ecology*, **24**, 4222–4237.

838 Sonesson C, Delorenzi M (2013) A comparison of methods for differential expression analysis  
839 of RNA-seq data. *BMC Bioinformatics*, **14**, 91.

840 Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of  
841 large phylogenies. *Bioinformatics (Oxford, England)*, **30**, 1312–3.

842 Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next generation.  
843 *Trends in ecology & evolution*, **25**, 705–12.

844 Staubach F, Teschke M, Voolstra CR, Wolf JBW, Tautz D (2010) A test of the neutral model of  
845 expression change in natural populations of house mouse subspecies. *Evolution*, **64**,  
846 549–560.

847 Stopková R, Stopka P, Janotová K, Jedelský PL (2007) Species-specific expression of major  
848 urinary proteins in the house mice (*Mus musculus musculus* and *Mus musculus*  
849 *domesticus*). *Journal of Chemical Ecology*, **33**, 861–869.

850 Stopková R, Vinkler D, Kuntová B *et al.* (2016) Mouse Lipocalins (MUP, OBP, LCN) Are Co-  
851 expressed in Tissues Involved in Chemical Communication. *Frontiers in Ecology and*  
852 *Evolution*, **4**, 47.

853 Turner L, Harr B (2014) *Genome-Wide Mapping In A House Mouse Hybrid Zone Reveals*  
854 *Hybrid Sterility Loci And Dobzhansky-Muller Interactions*. Cold Spring Harbor Labs  
855 Journals.

856 Turner LM, Schwahn DJ, Harr B (2012) Reduced male fertility is common but highly variable  
857 in form and severity in a natural house mouse hybrid zone. *Evolution*, **66**, 443–458.

858 Uebbing S, Künstner A, Mäkinen H *et al.* (2016) Divergence in gene expression within and  
859 between two closely related flycatcher species. *Molecular Ecology*, **25**, 2015–2028.

860 Utsumi M, Ohno K, Kawasaki Y *et al.* (1999) Expression of major urinary protein genes in the

861 nasal glands associated with general olfaction. *Journal of Neurobiology*, **39**, 227–236.

862 Woolstra C, Tautz D, Farbrother P, Eichinger L, Harr B (2007) Contrasting evolution of  
863 expression differences in the testis between species and subspecies of the house  
864 mouse. *Genome Research*, **17**, 42–49.

865 Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics.  
866 *Nature Reviews Genetics*, **10**, 57–63.

867 Wang J, Kalyan S, Steck N *et al.* (2015) Analysis of intestinal microbiota in hybrid house mice  
868 reveals evolutionary divergence in a vertebrate hologenome. *Nature Communications*,  
869 **6**, 6440.

870 Wilkinson GS, Breden F, Mank JE *et al.* (2015) The locus of sexual selection: moving sexual  
871 selection studies into the post-genomics era. *Journal of Evolutionary Biology*, **28**, 739–  
872 755.

873 Wittkopp PJ, Haerum BK, Clark AG (2008) Regulatory changes underlying expression  
874 differences within and between *Drosophila* species. *Nat Genet*, **40**, 346–350.

875 Wolf JBW, Bayer T, Haubord B *et al.* (2010) Nucleotide divergence vs. gene expression  
876 differentiation: comparative transcriptome sequencing in natural isolates from the  
877 carrion crow and its hybrid zone with the hooded crow. *Molecular Ecology*, **19**, 162–  
878 175.

879 Wynn EH, Sanchez-Andrade G, Carss KJ, Logan DW (2012) Genomic variation in the  
880 vomeronasal receptor gene repertoires of inbred mice. *Bmc Genomics*, **13**, 13.

881 Yang H, Zhang YP (2007) Genomic organization and sequence analysis of the vomeronasal  
882 receptor V2R genes in mouse genome. *Chinese Science Bulletin*, **52**, 336–342.

883 Young JM, Shykind BM, Lane RP *et al.* (2003) Odorant receptor expressed sequence tags  
884 demonstrate olfactory expression of over 400 genes, extensive alternate splicing and  
885 unequal expression levels. *Genome Biology*, **4**, 15.

886 Zhang XM, Firestein S (2002) The olfactory receptor gene superfamily of the mouse. *Nature*  
887 *Neuroscience*, **5**, 124–133.

888 Zhang XH, Marcucci F, Firestein S (2010) High-Throughput Microarray Detection of  
889 Vomeronasal Receptor Gene Expression in Rodents. *Frontiers in Neurosciences*, **4**, 164.

890 Zhang XH, Zhang XM, Firestein S (2007) Comparative genomics of odorant and pheromone  
891 receptor genes in rodents. *Genomics*, **89**, 441–450.

892

893

894 DATA ACCESSIBILITY

895 Raw reads have been archived in the NCBI SRA public database under the bioproject  
896 identifier PRJNA356041 and through the biosample identifiers SAMN06212561 to  
897 SAMN06212563 and SAMN06212565 to SAMN06212572. Expression values (raw and  
898 normalised counts) for each gene in each sample are provided in Table S3.

899

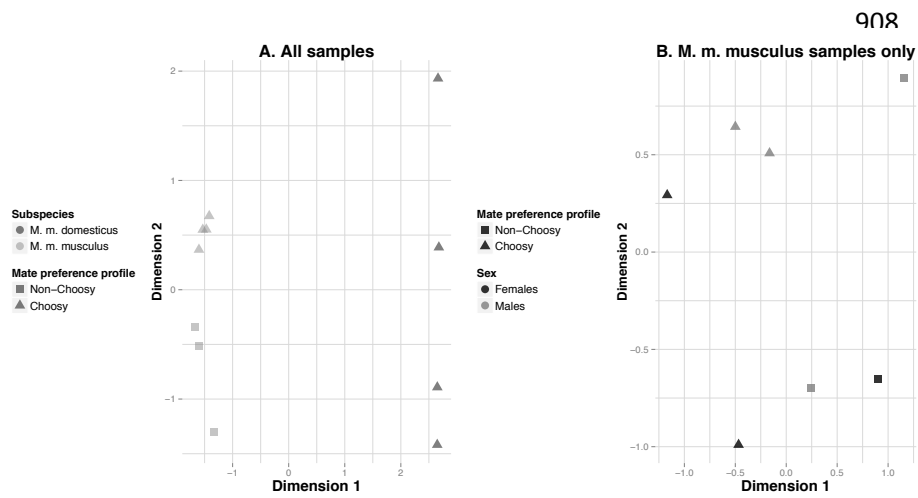
900 Author contributions

901 C.S., P.B. and G.G. conceived the experiment. C.S., G.G. and P.C. conducted fieldwork and  
902 dissections. P.C. and D.S. performed RNA sample and library preparation and sequencing.  
903 E.L. and S.T. analysed the data. C.S. wrote the manuscript and E.L., S.T. and G.G. provided  
904 comments and revisions on it.

905

906 MAIN FIGURES

907



915

916

917 **Figure 1:** Multi-Dimensional Scaling (MDS) plots indicating variation in gene expression

918 among (A) all *M. m. musculus* and *M. m. domesticus* samples and (B) *M. m. musculus*

919 samples only.

920

921

922

923

924

925

926

927

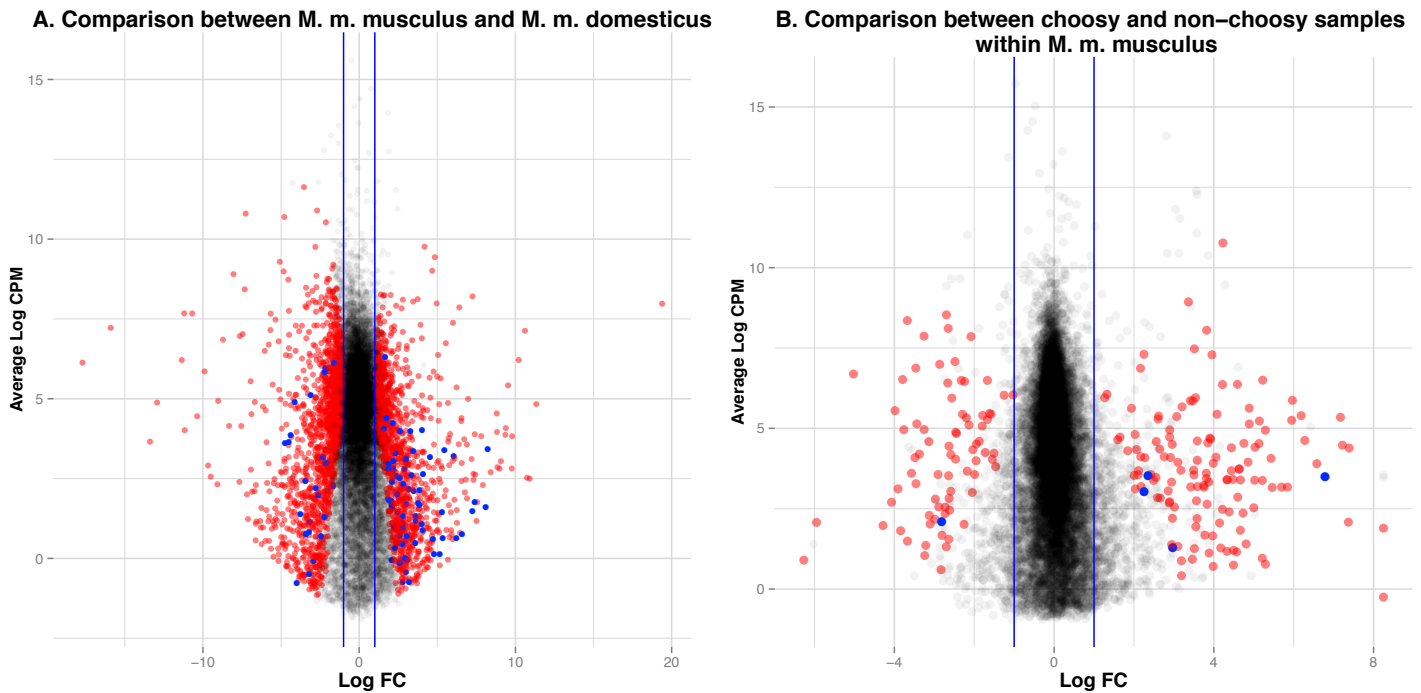
928

929

930

931

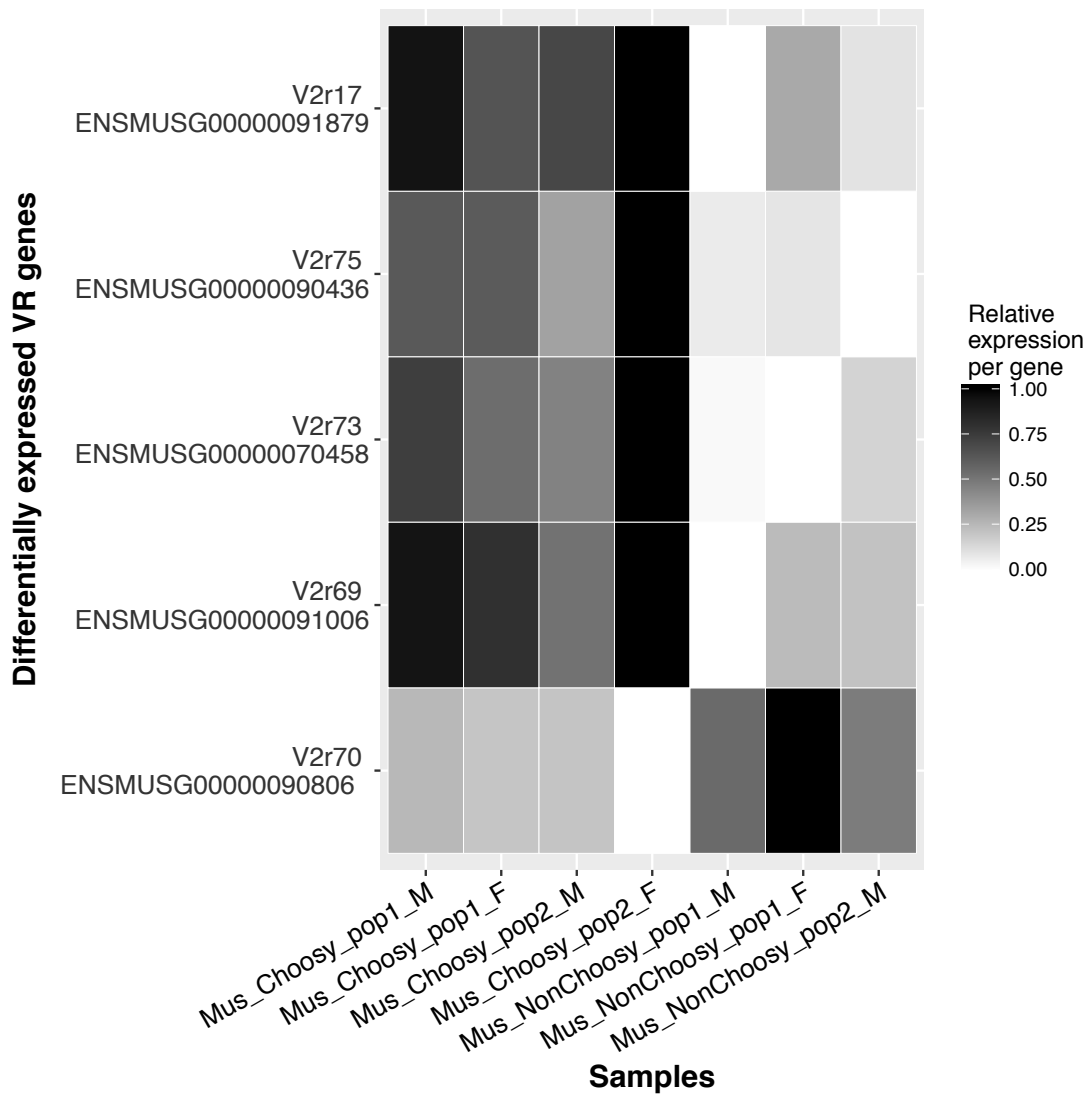




932

933 **Figure 2:** Smear plots indicating significantly differentially expressed genes (red dots) and  
 934 vomeronasal receptors (blue dots) between (A) *M. m. musculus* 'Choosy' and *M. m.*  
 935 *domesticus* 'Choosy' samples; (B) 'Choosy' and 'Non-Choosy' samples within *M. m. musculus*.  
 936 Non-differentially expressed genes are shown as black dots. The blue vertical lines indicate 1  
 937 log fold change. The x-axis indicates genes that are up regulated in each sample (plot A: dots  
 938 on the right and on the left corresponding to genes up regulated in *M. m. musculus* and in  
 939 *M. m. domesticus* respectively; plot B: dots on the right and on the left corresponding to  
 940 genes up regulated in 'Choosy' *M. m. musculus* and in 'Non-Choosy' *M. m. musculus*  
 941 respectively). The y-axis indicates the level of average expression of the gene (highly  
 942 expressed with higher values). CPM: counts per million; FC: fold change.

943



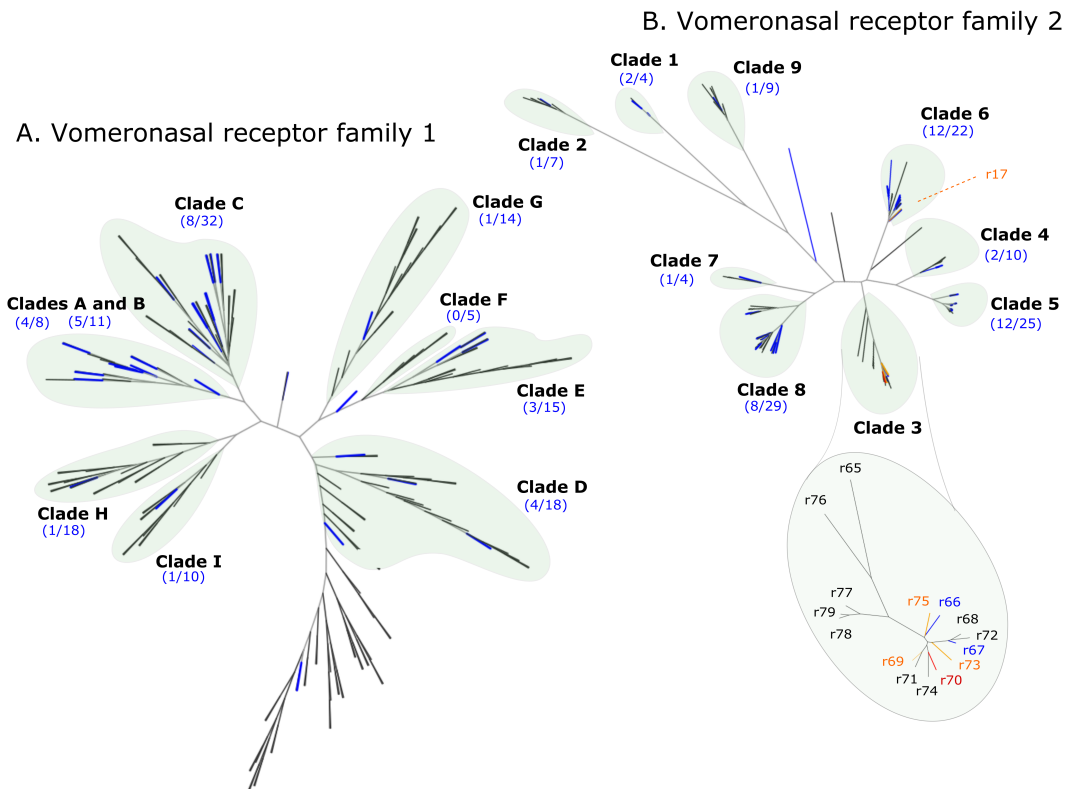
945

946

947 **Figure 3:** Heatmap plot of differentially expressed vomeronasal receptor (VR) genes in the  
 948 comparison between 'Choosy' and 'Non-Choosy' *M. m. musculus* samples. The values are  
 949 presented in log fold change (FC) per gene, in which the sample with the lowest expression  
 950 value corresponds to 0.

951

952



954

955

956 **Figure 4:** Phylogenetic trees of the two families of vomeronasal receptors (VR) (A: family 1;  
 957 B: family 2). Branches coloured in blue indicate differentially expressed VR between *M. m.*  
 958 *musculus* and *M. m. domesticus*. The number of differentially expressed VR genes between  
 959 the two subspecies and the total number of VR genes in each subfamily (clade) are  
 960 presented between brackets (e.g. 8/29). Branches coloured in orange indicate differentially  
 961 expressed VR between ‘Choosy’ and ‘Non-Choosy’ samples within *M. m. musculus*. Branches  
 962 coloured in red indicate VR genes with significant differential expression in both  
 963 comparisons. In family 2, clade 3 is shown in detail with receptor IDs (MGI nomenclature,  
 964 see Methods).

965

966

## Supplemental Information for:

967

Do changes in gene expression contribute to sexual isolation and reinforcement in the house mouse?

969

970

971 Etienne Loire, Sergio Tusso, Pierre Caminade, Dany Severac, Pierre Boursot, Guila Ganem and

972

Carole M. Smadja

973

974

975

### Table of Contents:

976

List of supplementary figures and tables	Page 2
Supplementary text	Page 3
Figure S1	Page 4
Figure S2	Page 5
Figure S3	Page 6
Figure S4	Pages 7-11

977

978

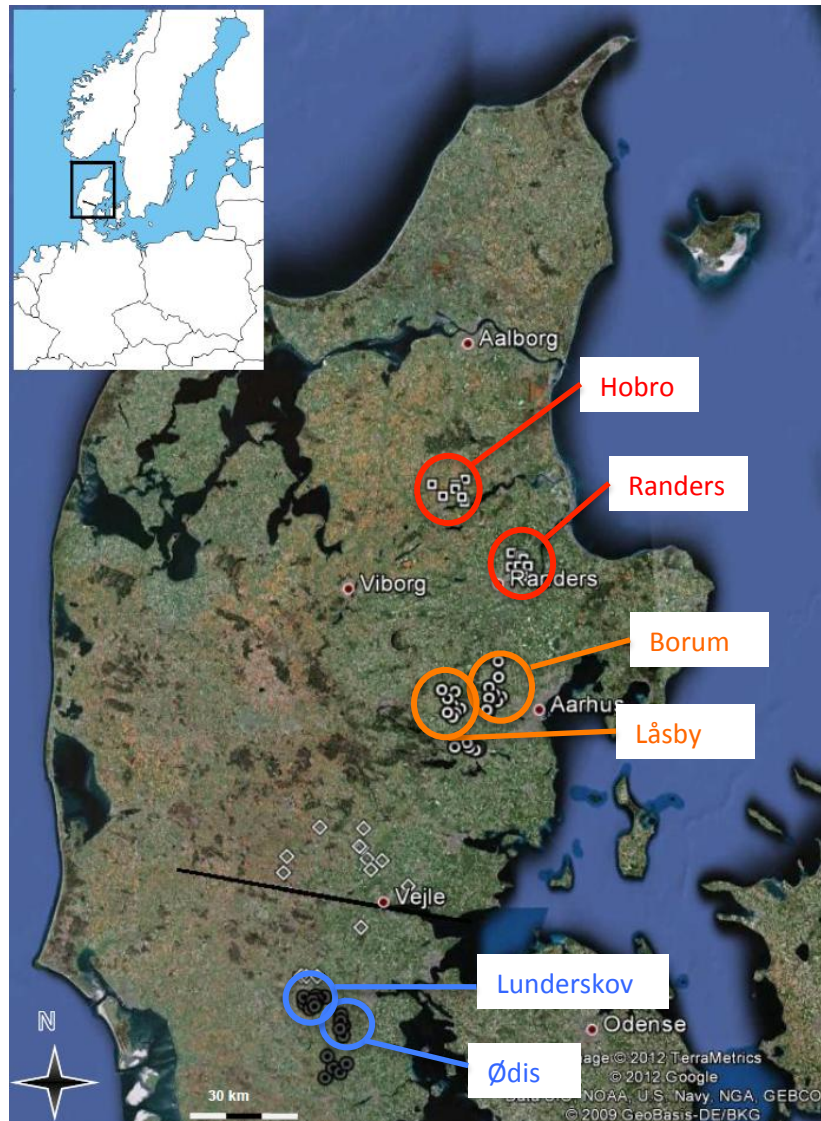
979 LIST OF SUPPLEMENTARY FIGURES AND TABLES:  
980  
981 **Table S1:** Sample description (separated Excel file)  
982 **Table S2:** General mapping and gene expression statistics (separated Excel file)  
983 **Table S3:** Raw and normalised (CPM) counts for all genes and all samples (separated Excel  
984 file, sheet 1: ALL\_GENES\_raw\_counts; sheet 2: ALL\_GENES\_normalised\_counts, sheet 3:  
985 OR\_normalised\_counts; sheet 4: VR\_normalised\_counts; sheet 5: MUP\_normalised\_counts;  
986 sheet 6: OBP\_normalised\_counts)  
987 **Table S4:** Variation in mean count and in expressed gene number among samples (separated  
988 Excel file, sheet 1: Table S4a: number and percentage of expressed genes per sample; sheet  
989 2: Table S4b: average counts per sample)  
990 **Table S5:** List of differentially expressed genes (separated Excel file, sheet 1: Table S5a:  
991 Differentially expressed genes between "Choosy" *Mus musculus musculus* and "Choosy" *M.*  
992 *m. domesticus*; sheet 2: Table S5b: Differentially expressed genes between "Choosy" and  
993 "Non-Choosy" samples within *Mus musculus musculus*)  
994 **Table S6:** Functional enrichment results (separated Excel file, sheet 1: Table S6a: enriched  
995 functional categories in differentially expressed genes between 'musculus Choosy' and  
996 'domesticus Choosy' samples; sheet 2: Table S6b: enriched functional categories in  
997 differentially expressed genes between 'musculus Choosy' and 'musculus Non-Choosy'  
998 samples)  
999  
1000 **Figure S1:** Map indicating the geographical origins of the mice used in this study.  
1001 **Figure S2:** Violin plots of log-transformed normalised gene coverage.  
1002 **Figure S3:** Distribution of phylogenetic distances among random sampling of five  
1003 vomeronasal receptor genes.  
1004 **Figure S4:** Coverage per position (bp) for the five differentially expressed VR genes in the  
1005 comparison between 'Choosy' and 'Non-Choosy' *M. m. musculus* samples.  
1006  
1007

1008 **Supplementary information**

1009 Discussion:

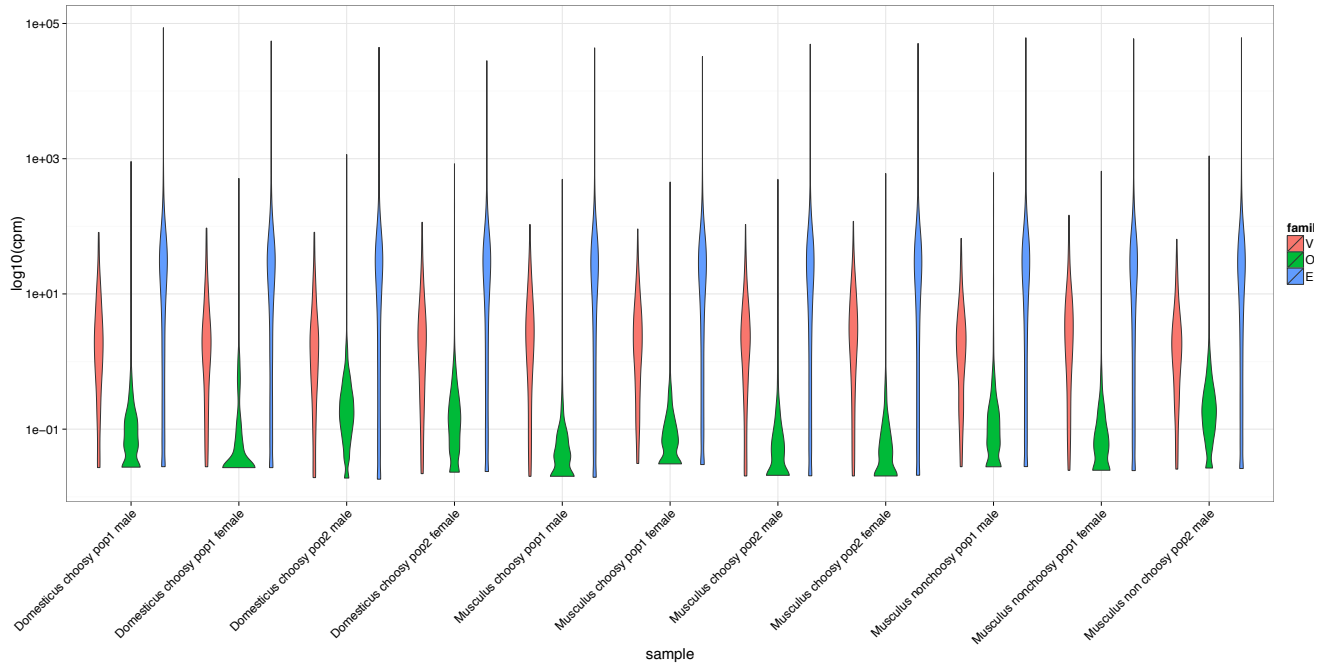
1010 *Robustness of differential expression inferences*

1011 RNA-seq on pooled samples is a cost-effective strategy widely used to assess variation in  
1012 gene expression among experimental conditions or biological groups while producing  
1013 transcriptome-wide data (reviewed in Todd *et al.* 2016). This pooling design offers an  
1014 advantage by increasing sample size into the analysis, and when working with samples that  
1015 may exhibit a large amount of biological variation, analysing several sample pools, rather  
1016 than a few single samples per condition, also lessens the impact of single aberrant samples  
1017 (Todd *et al.* 2016). Although sample pooling can theoretically result in lower precision in  
1018 gene expression estimation (for lowly expressed transcripts) and increased false positive  
1019 rates in differential expression (DE) analyses as compared to individual sequencing  
1020 (Rajkumar *et al.* 2015), this limitation is overcome when the number of individuals per pool  
1021 is sufficiently large ( $n > 5$ ) and the number of biological replicates per group is maximized  
1022 (Todd *et al.* 2016). By using eight individuals per pool and from three to four biological  
1023 replicates per group for DE analyses, we placed ourselves in favourable conditions to reliably  
1024 identify changes in gene expression between groups with a pooling design. To minimize  
1025 further the risk of false positives in DE analyses, we also applied a stringent false discovery  
1026 correction and selected the method for DE detection (*edgeR*) with the highest sensitivity and  
1027 specificity compared to other methods (e.g. *DESeq2*, *Cuffdiff2*) (Rajkumar *et al.* 2015).  
1028 Altogether, the stringent DE analysis and the consistency of differential expression among  
1029 biological replicates strongly support the idea of divergence in gene expression between  
1030 'Choosy' and 'Non-Choosy' *M. m. musculus* samples.



1031  
 1032 **Figure S1:** Map indicating the geographical origins of the mice used in this study (adapted  
 1033 from Latour et al. 2014).  
 1034 Trapping sites in Jutland (Denmark) used in Latour et al. 2014 (the bold black line represents  
 1035 the genetic centre of the hybrid zone as defined in Raufaste et al 2005).  
 1036 The colored circles indicate the trapping sites defining the populations analysed in this study  
 1037 (red: “Non-Choosy *M. m. musculus* populations”; orange : “Choosy *M. m. musculus*  
 1038 populations”; blue: “Choosy *M. m. domesticus* populations”

1039  
1040  
1041  
1042

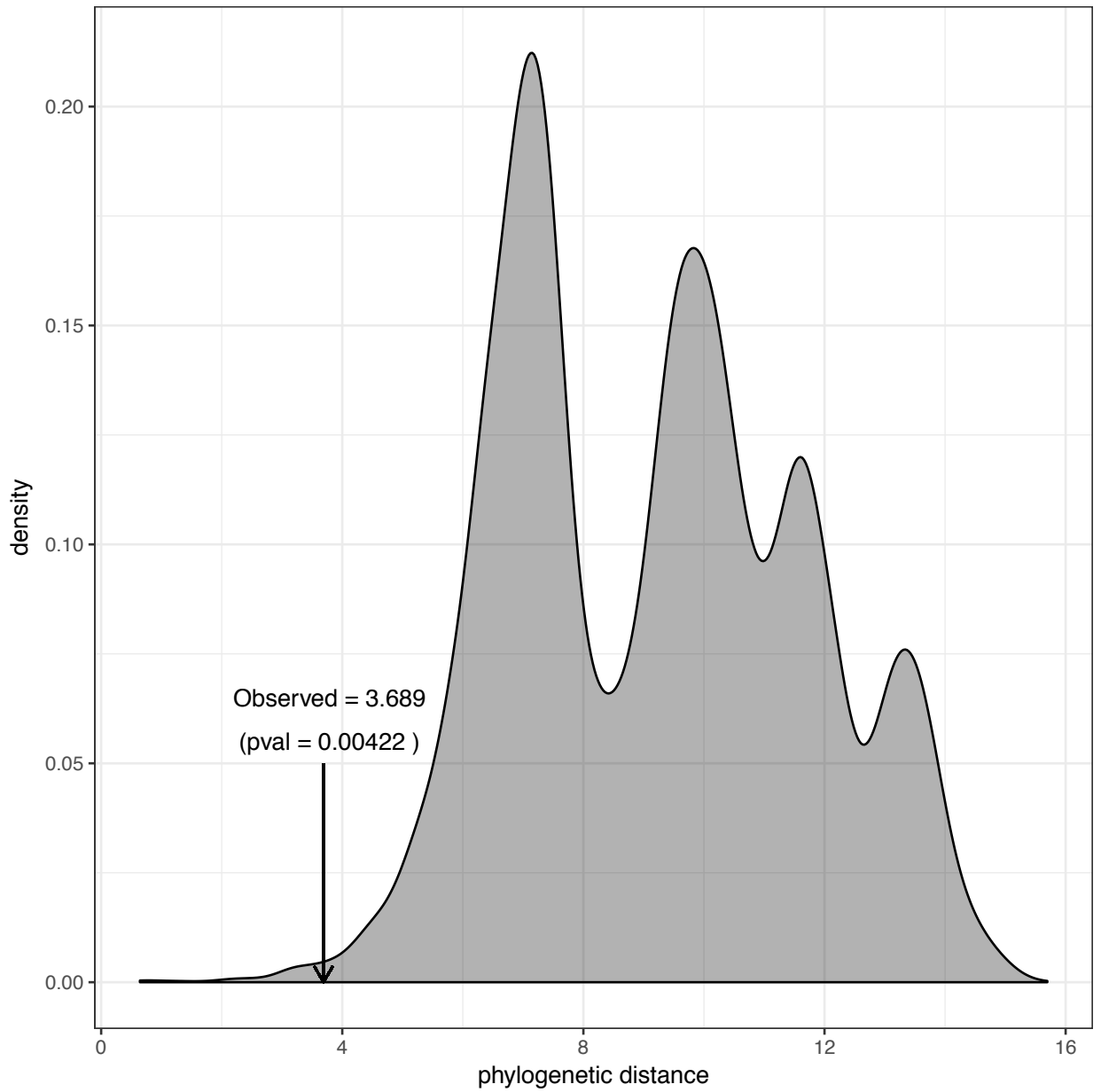


1043  
1044  
1045  
1046  
1047  
1048  
1049

**Figure S2:** Violin plots of log-transformed normalised gene coverage (CPM) per population for vomeronasal receptor (VR) (red), olfactory receptor (OR) (green) and all other categories of genes (blue).



1050  
1051  
1052



1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061

**Figure S3:** Distribution of phylogenetic distances among random sampling of five vomeronasal receptor genes. The arrow represents the observed distance between the five VRs found differentially expressed between 'Choosy' and 'Non-Choosy' samples, and its probability.

1062

1063

1064

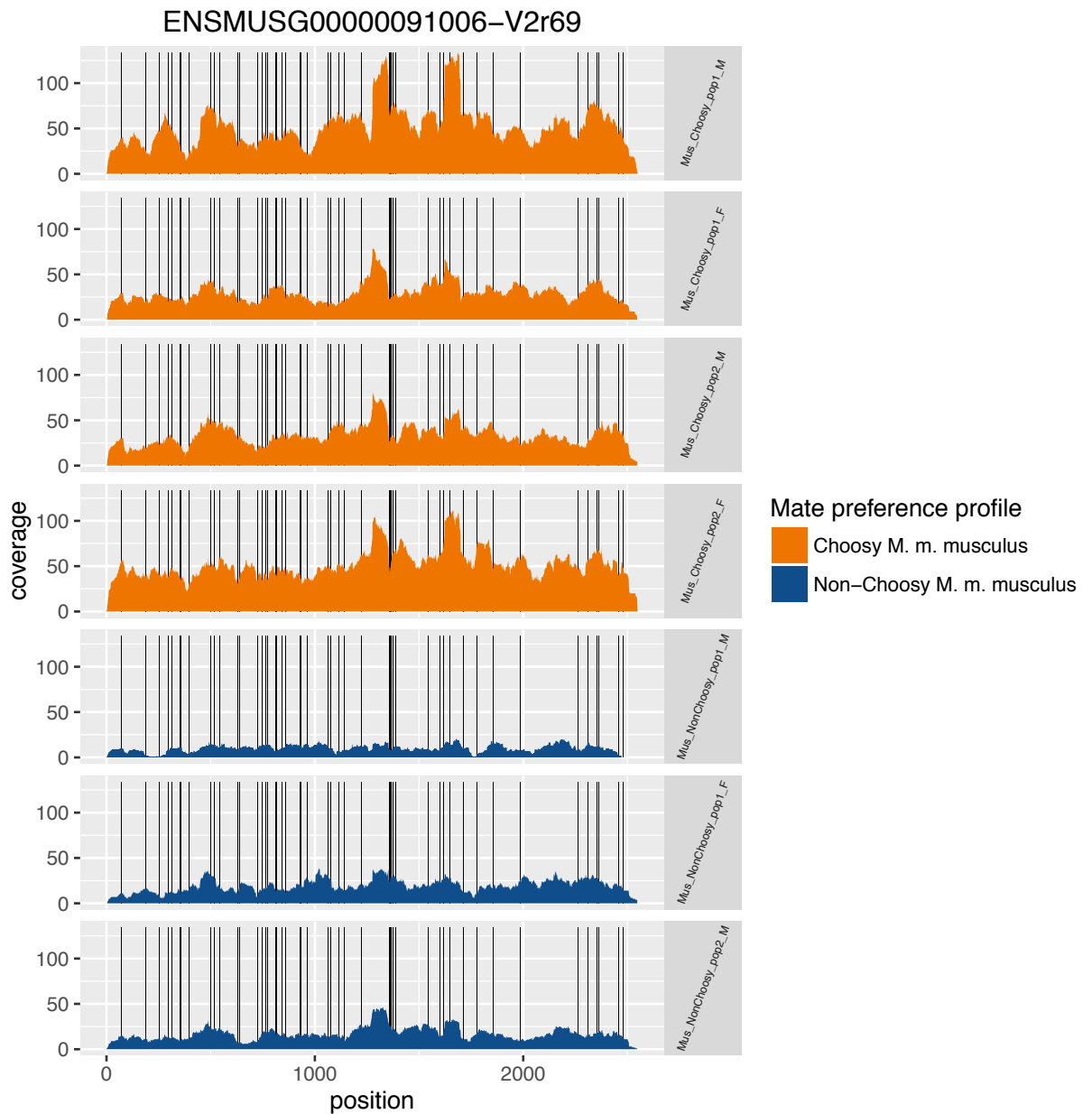
1065

1066

1067

1068

**Figure S4:** Coverage per position (bp) for the five differentially expressed VR genes in the comparison between 'Choosy' and 'Non-Choosy' *M. m. musculus* samples. Vertical lines represent variable sites of the gene.



1069

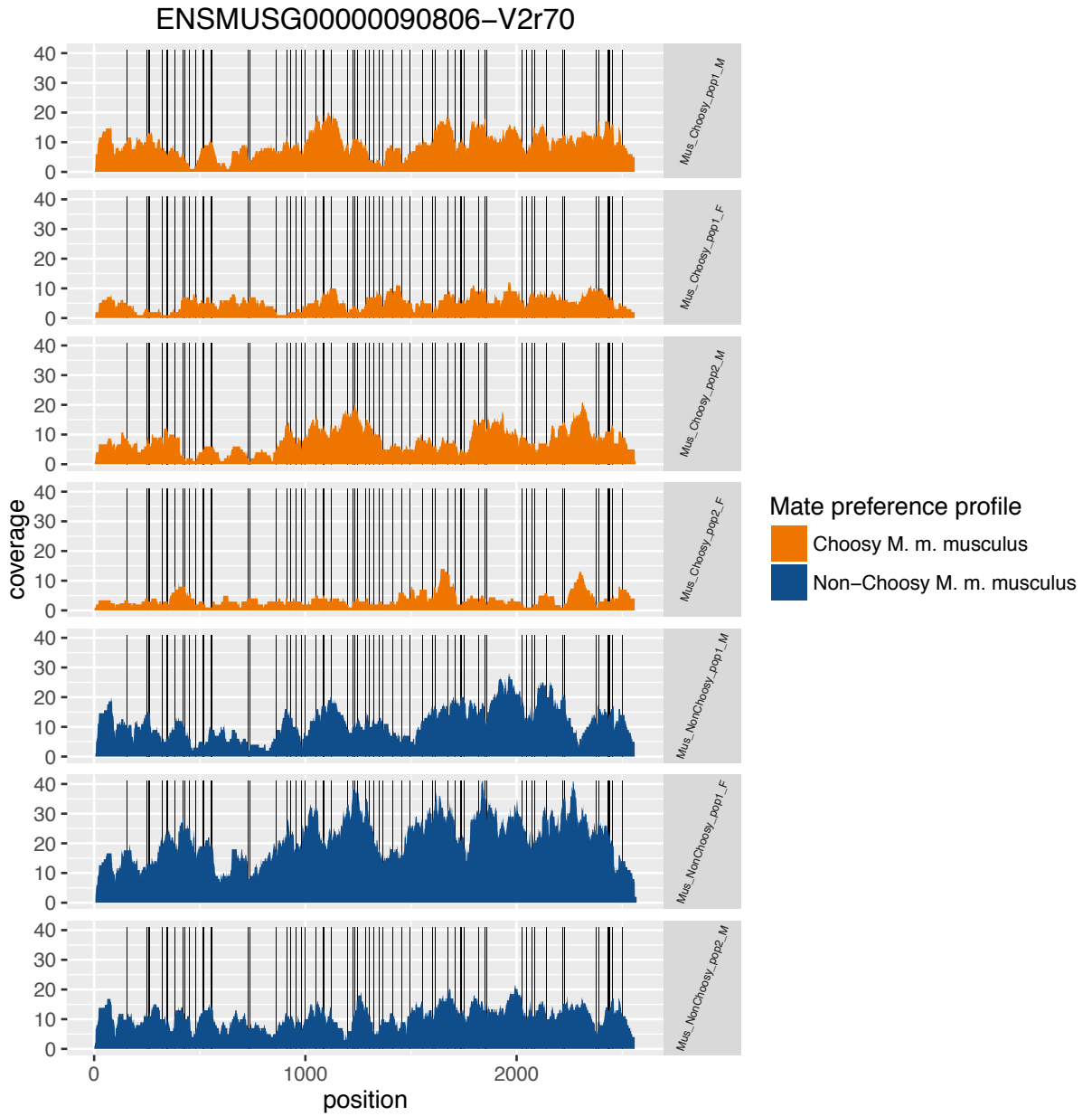
1070

1071

1072

1073

1074



1075

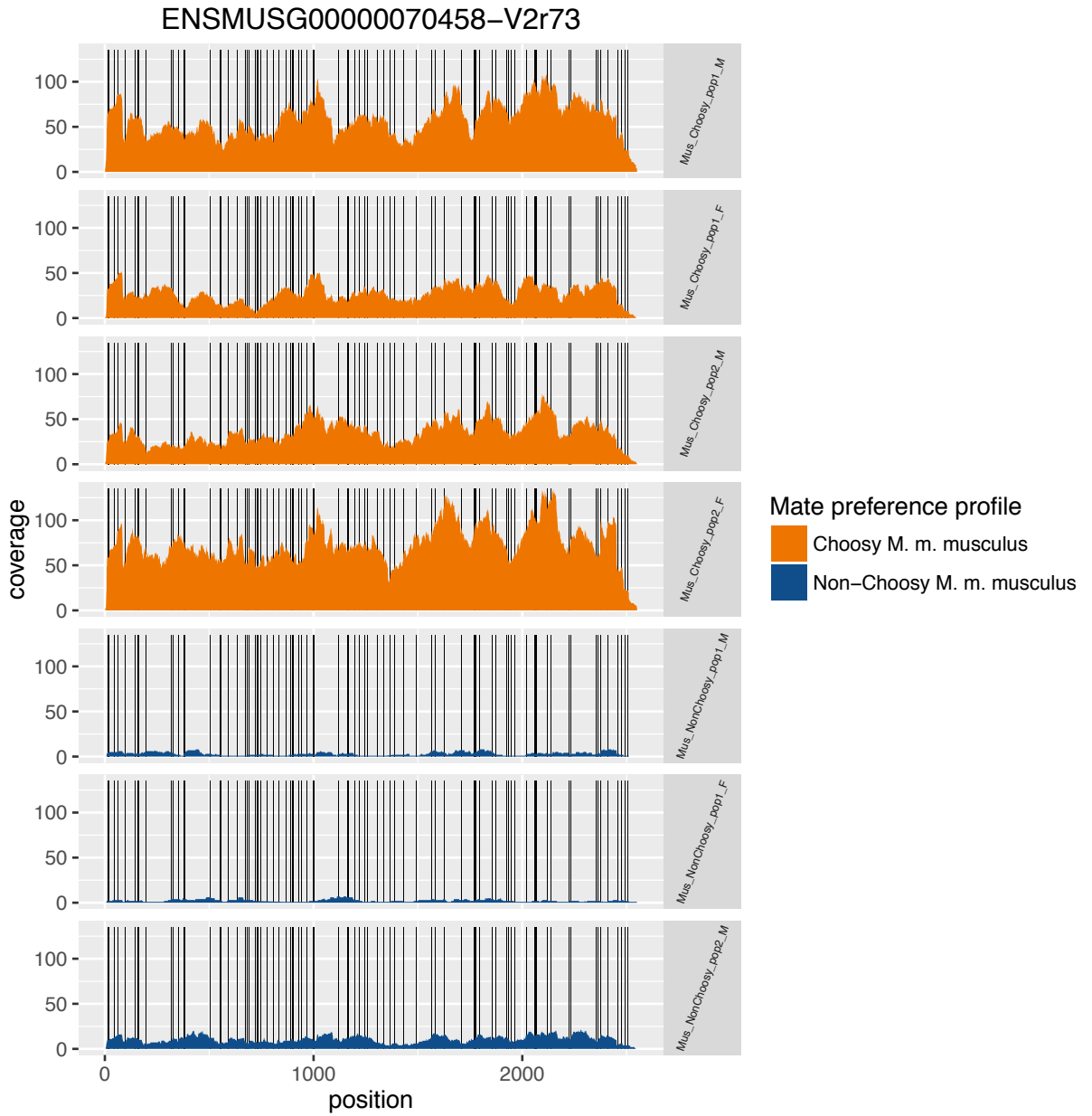
1076

1077

1078

1079

1080



1081

1082

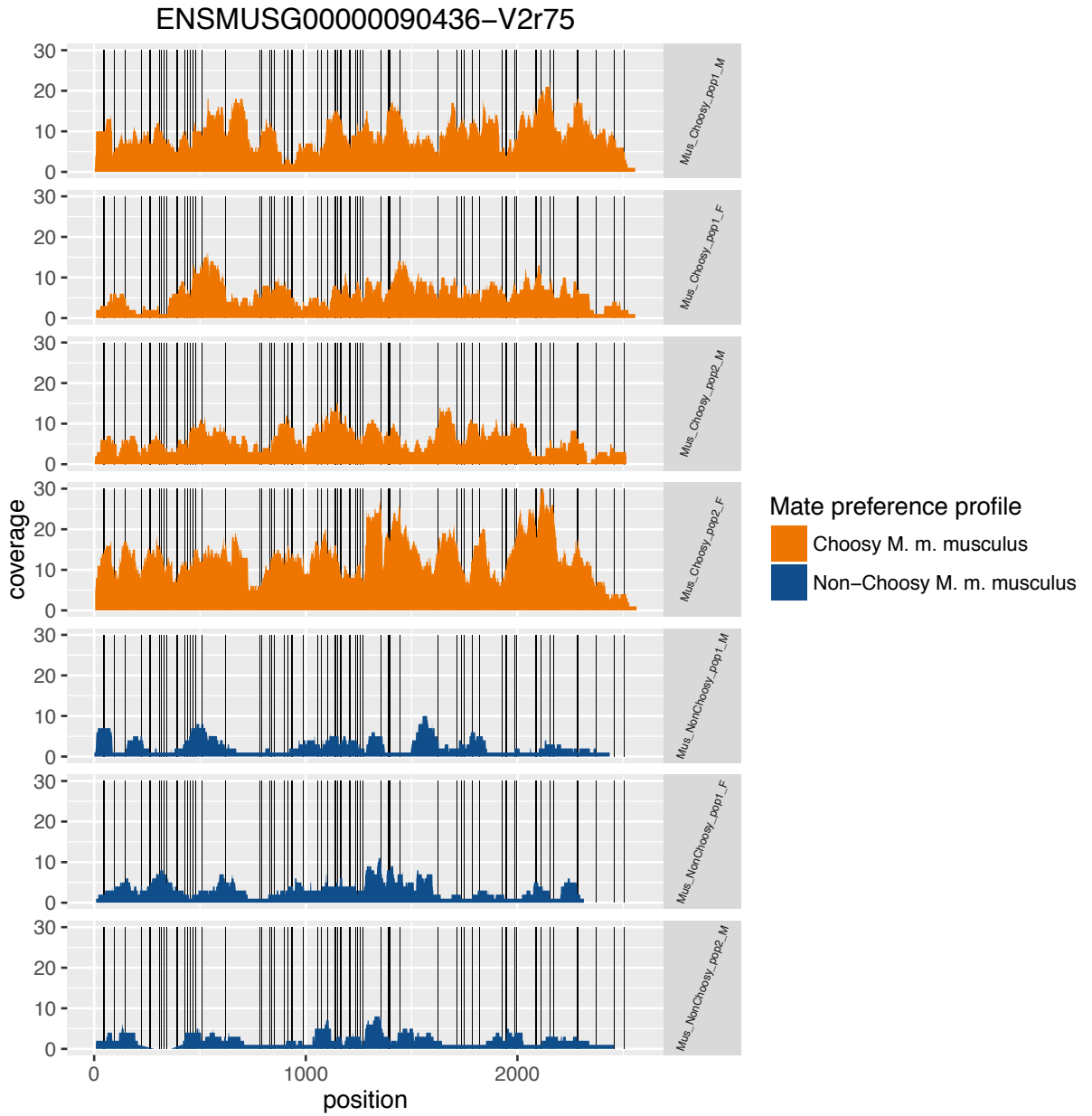
1083

1084

1085

1086

1087



1088

1089

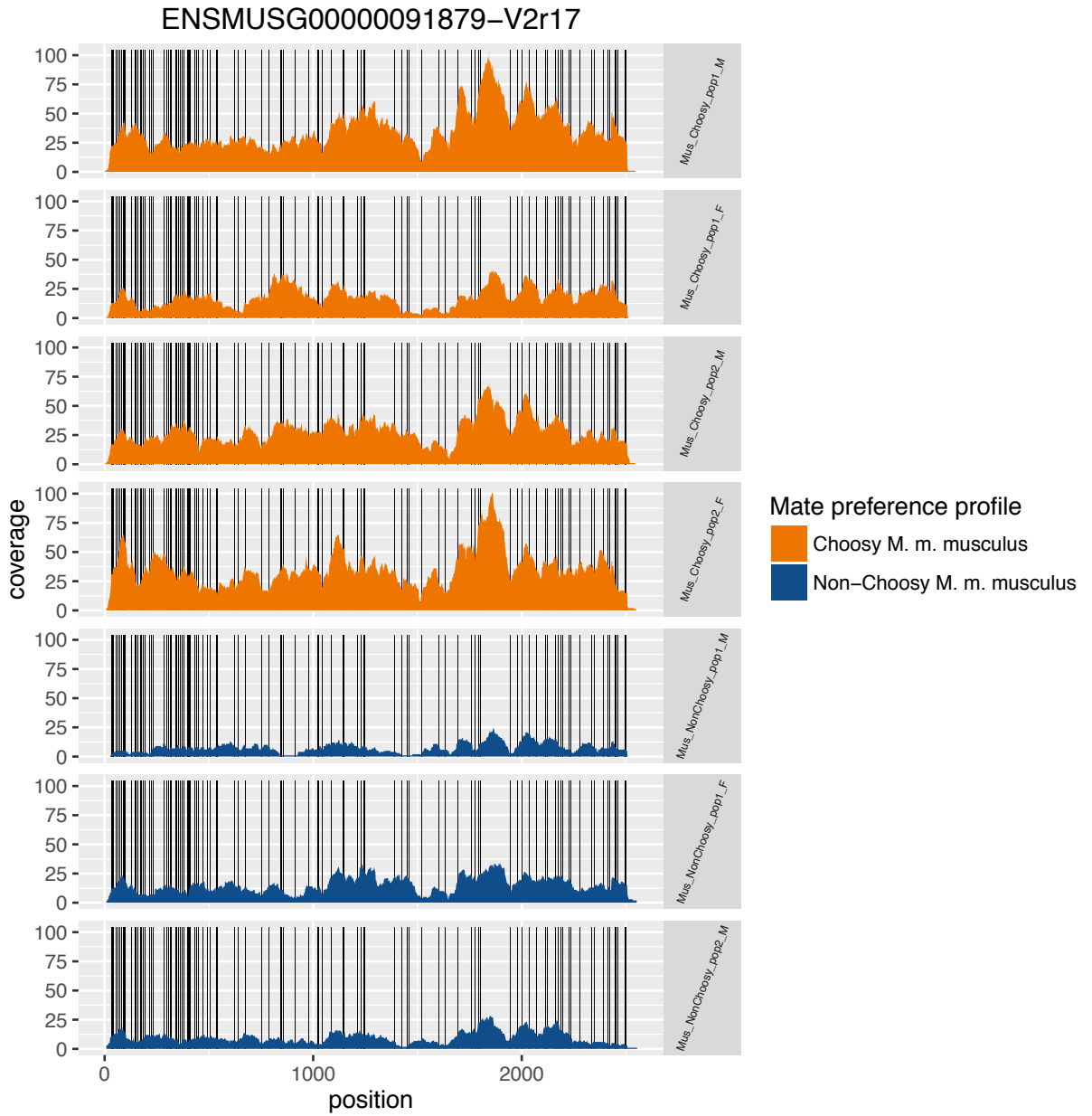
1090

1091

1092

1093

1094



1095

1096