Identification of genomic regions associated with tameness using selective breeding of a novel outbred mouse stock

Matsumoto, Yuki

Doctor of Philosophy

Department of Genetics

School of Life Science

SOKENDAI

Abstract

Tameness is one of the major behavioral factor for animal domestication, and the molecular genetic mechanism should be shared between each domesticated animal species. To identify the genetic region associated with tameness in mammalian species, I applied selective breeding for active tameness, one aspect of tameness involves motivation to approach humans, and following selection mapping using the mouse as a model organism. Given that the prerequisite for successful selective breeding is high genetic variation in the target population, my collaborator and I established and utilized a novel resource, wild-derived heterogeneous stock mice, from eight wild mouse strains. As a result of selective breeding of the wild-derived heterogeneous stock, the level of contacting which strongly associated with active tameness increased through the generations. Applying selection mapping to the selected population using a simulation based on a non-selection model and over 20 thousand single-nucleotide polymorphisms, I found a genomic signature of selection on the Chromosome 11. Following association analysis on the Chromosome 11 identified two-closely linked loci associated with active tameness (Active tame related (ATR) region). Further I used comparative genetic analyses and confirmed whether or not the ATRs overlap the locus to rat tameness QTL as well as the dog genomic regions which are known to be a region selected during the dog domestication. I found an overlap region (0.61Mb) within the ATR and the result suggested that the region could be associated with tameness in the mammalian species.

General introduction

Tameness is one of the major element in animal domestication (Price 2002, Grandin and Deesing 2014) and it has important role especially in early stage of the domestication (Wilkins et al. 2014). Although tameness is difficult to define completely, tameness could involve many factors such as anxiety, fear, novelty-seeking (Goto et al. 2013), Price (2002) proposed that tameness can be divided into two potential components, motivation to approach humans (active tameness) and reluctance to avoid them (passive tameness).

Several studies on tameness in experimental condition indicated that tameness is affected by genetic factor in foxes (*Vulpes vulpes*: Kukekova et al. 2011), rats (*Rattus norvegicus*: Albert et al. 2009), mice (*Mus musculus*: Goto et al. 2013), and red junglefowl (*Gallus gallus*: Agnvall et al. 2012). In addition, many animals showing tameness suggesting the genetic component underling tameness should be share among different animal species, and the number of causative "master" genetic factor could be a few (Grandin and Deesing 2014). Identification of master genetic factor underling tameness contribute to understand how animals were domesticated, especially in the early stage of the domestication.

Previously, genetic analyses using genome-wide association studies (GWAS) and quantitative trait locus (QTL) mapping were conducted to identify the genetic loci associated with tameness in rats and foxes (Albert et al. 2009; Kukekova et al. 2011). Recently, selection mapping studies, a method identifies loci using populations that have been subjected to natural/artificial selection with recombination, also have tried to identify domestication gene(s) including tame related loci in the domesticated animals

(Carneiro et al. 2014; Montague et al. 2014). However, little evidence has been obtained so far due to difficulties of controlling genetic background and obtaining poor information on the genetic markers for the subject species. In contrast to these animals, mouse is a useful model organism for conducting GWAS, QTL mapping, and selection mapping to identify genetic basses of tameness. Several advantages of mice involve easily control of genetic and environmental heterogeneity affecting genetic mapping analysis, and accessibility to a large amount of data in the current databases.

Here I conducted a study comprised from three processes to identify genetic loci associated with tameness in mammalian species. I chose mouse as a model to identify the genetic region associated with tameness, and after the identification I performed comparative genetic analysis to reveal the genetic regions potentially influencing tameness in the other mammals. First of all, eight wild mouse strains were used to establish a novel outbred mouse stock, wild-derived heterogeneous stock (WHS), and I conducted selective breeding for "contacting," which is strongly associated with active tameness (Goto et al. 2013), using the WHS (Chapter 1). The selected and non-selected populations for contacting should be a valuable resource for further study in mice. Second, I performed selection and association mapping to identify selected loci associated with contacting (Chapter 2). Finally, I conducted comparative analysis of the mapped loci with two mammals, rats and dogs, to study the existence of a shared genetic basis of tameness (Chapter 3).

CHAPTER 1: Establishing high tame mice via a selective breeding

1.1. Introduction

In contrast to existing domesticated and wild animal populations, mouse is a useful model organism to conduct genetic mapping. Heterogeneous stock (HS), an outbred offspring descended from eight founder strains (Flint and Eskin 2012), is a mouse genetic resource for genome-wide association studies (GWAS) to detect genes affecting complex traits (Valdar et al. 2006). Two HS, Boulder HS (BHS) and Northport HS (NHS), have been used for genetic mapping experiments (Yalcin and Flint 2012). BHS is derived from eight laboratory strains (A/J, AKR, BALB/c, C3H, C57BL/6, DBA/2, I, and RIII) and has been bred over 60 generations (Yalcin and Flint 2012). NHS is also derived from eight laboratory strains (A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J, and LP/J), and has been bred for 40 generations (Chia et al. 2005). In recent years, Collaborative Cross, a large panel of recombinant inbred lines derived from five laboratory strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/LtJ, and NZO/HiLtJ) and three wild strains (CAST/EiJ, PWK/PhJ, and WSB/EiJ), has been developed (Threadgill et al. 2002). By using progenitor mice at the early stage of inbreeding in the establishment of Collaborative Cross, Diversity Outbred (DO) was developed by a randomized outbreeding strategy with 175 breeding pairs (Svenson et al. 2012). Given that the existence of genetic diversity in these resources is one of the most important prerequisites for genetic mapping with HS (Flint and Eskin 2012), limited heterogeneity in the laboratory strains used as all founder strains of BHS and NHS could be a problem for conducting genetic studies. HS can also be used to conduct selective breeding for behavioral traits (DeFries et al. 1978; Belknap et al. 1983;

Grahame 1999; Hitzemann et al. 2009). For selective breeding, genetic diversity of the resource is important also (McClearn et al. 1970; Zombeck et al. 2011). Therefore, incorporating more wild strains that have greater genetic diversity than laboratory strains is advantageous for mapping experiments and selective breeding.

Here I chose eight wild mouse strains to establish wild-derived heterogeneous stock (WHS) and revealed the genetic characteristics of the strains by using genome-wide SNP data. I established WHS through mixing genomes of the eight strains and conducted selective breeding for "contacting," which is strongly associated with active tameness (Goto et al. 2013), using the WHS. Next I analyzed haplotype contribution in the stock to characterize heterogeneity. In this chapter, I showed characteristics of selected and non-selected populations which could be valuable resource for further study to identify genetic loci associated with tameness (Chapter 2 and Chapter 3).

1.2. Methods

1.2.1. Animals

All mice were maintained in accordance with NIG guidelines and all procedures were carried out with approval (No. 26-9) from the Committee for Animal Care and Use of the National Institute of Genetics (NIG). Mice were bred and kept under specific-pathogen-free conditions at the NIG. Availabilities of food and water was free under a 12/12-h light/dark cycle in a temperature-controlled room (23 ± 2°C). To establish WHS, my collaborator and I used eight wild-derived strains, BFM/2, PGN2, HMI, BLG2, NJL, KJR, CHD, and MSM (Table 1). These eight strains were originated

from wild caught mice in the all over the world (Koide et al. 2000). In order to reduce pain, tweezers covered with silicon tubing were used to catch the mouse tail when I exchange the cage and perform behavioral tests described later.

1.2.2. Genetic structure of WHS founders

To characterize the genetic component of WHS with comparing other heterogeneous stock mice, I analyzed single nucleotide polymorphisms (SNP) in the eight strains. To genotype the eight strains I used the 77K MegaMUGA array (Geneseek, Lincoln, NE, USA), which is an Illumina based SNP genotyping array (Welsh et al. 2012). Further, I obtained additional data for other mouse strains from available SNP data of MegaMUGA for 45 mouse strains. These strains included founder strains of other heterogeneous stock (BHS, NHS and DO) from UNC Systems Genetics at the University of North Carolina (http://csbio.unc.edu/CCstatus/index.py?run=GeneseekMM). I used SNP data from total 53 strains and the data was controlled by PLINK v. 1.9 (Chang et al. 2015). I removed SNPs that were identical among all strains and missing sites and then 10,598 SNPs were remained after quality control.

Next the relationships among the founder strains of WHS and other inbred mouse strains was evaluated by using Neighbor-joining method (Saitou and Nei 1987) with the p-distance. Then evolutionary history among the founder strains was inferred using MEGA version 6.06 (Tamura et al. 2013). The 1,000 bootstrap tests were conducted.

1.2.3. Establishing wild-derived heterogeneous stock

In order to establish WHS, I mated eight wild strains BFM/2, PGN2, HMI, BLG2, NJL, KJR, CHD, and MSM. A male mouse of each strain was mated with a female of

another strain t generation zero (G_0). Then I obtained pups of each mated pairs at G_1. The circular rotation rule (Koide et al. 2012) was used to mate following generation. At the G_2 generation, the number of pairs were expanded from 8 to 16 and mated again following the rotation rule. To avoid losing eight different genomes, the populations were expanded into two populations. As the size of heterogeneous stock with 16 pairs is small compared with other stocks, keeping two different population for selected and control may help to maintain the original alleles in each genetic loci. Mating pairs were made following the random mating rule while avoiding intercrossing from the G_3 generation. Then, 16 pairs in each population were mated in this study. In case of insufficient of offspring to a mating pair, I used substitute mice from the progeny from other pairs within the same population, which is also appropriate to avoid inbreeding. At the G_3 generation, the genomes of all of the eight strains were mixed randomly in each mouse. Then high levels of genetic diversity could be promising than it in the eight founder strains.

1.2.4. Selective breeding for active tameness

It is considered that currently available domesticated mouse strains had predominantly been selected for reluctance to avoid humans during their domestication, but not for motivation to approach humans (Goto et al. 2013). Few mouse strain showing high active tameness is exist. The establishment of a line selected for active tameness will be a crucial step for understanding of tameness in more detail.

I conducted breeding for two groups of selection and non-selection, as control. Each population was split into another two populations, selected and non-selected, from G_5 for the following reasons. First, the analyses to detect a selective sweep might be

affected by genetic drift. To distinguish a selective sweep against the potentially obscuring factor of genetic drift, I attempted to reproduce the results in duplicated groups. The second reason for establishing selected and non-selected populations was to increase the possibility of detecting selective sweep regions as each population consisted of a relatively small number of mice. The first two lines were selection 1 (S1) and control 1 (C1), and the additional two populations after splitting were selection 2 (S2) and control 2 (C2), which were derived from C1 and S1, respectively. Each line was kept with 14-16 breeding pairs during G_3 to G_{12}.

I applied following selection criteria only to the selected populations. For each mating of mice, I chose the mice based on highest scores of the behavioral indices, contacting and heading toward human hand, in the active tame test. Each offspring was ranked by contacting score within their family and sex. When the highest contacting scores between two mice were equal, I chose the mouse who exhibited a higher heading score for the next mating. In order to avoid inbreeding and environmental effect of the family (Lynch 1980), I did not perform crosses within each family. Each selected male and female from one family was subjected to mate to one from another family, to obtain the next generation.

I used random mating for control populations with avoiding inbreeding. A mouse for each male and female in each family was randomly chosen within the same sexes, and these were subjected to mate with a similarly chosen female and male from another family.

1.2.5. Haplotype inference of founder strains

In order to uncover the genetic contribution of eight founder strains to newly established WHS mouse population, I used inferring haplotypes for each local genomic

region in the autosomes. Here I genotyped 32 WHS mice from four populations at G_{12} and eight founder strains used by the 144K GigaMUGA SNP array (Geneseek) (Didion et al. 2014). After calling SNP, following quality control procedure was performed by using PLINK v. 1.9 (Chang et al. 2015). First I removed SNPs with identical nucleotides among the eight founder strains and those for which 1% or more of the data were missing. It is considered that the recombination rate of sex chromosomes was differ to some of the autosomes. The difference could affect the results of subsequent analyses. Thus the SNPs located on the X and Y chromosomes were removed. Finally, 52,135 SNPs were remained in 128 WHS mice and eight founder strains. The number of one strain specific SNP was 20,530 (39.4%) (Table 4).

The haplotypes of WHS founder strains were inferred by using Beagle v. 4.0 (Browning and Browning 2007) and RFMix v. 1.5.4, which is software for local-ancestry inference by a discriminative approach using random forest (Maples et al. 2013). Using the inferred data, I calculated the genetic contribution of each founder strain in each SNP site.

1.2. 6. Behavioral assay

My collaborator and I performed three different tame tests, active tame test, passive tame test, and stay-on-hand test, as described in the previous work (Goto et al. 2013). Total tested mice were over 300 mice of 6 weeks-of-age for each generation. These tests were implemented in an open-field apparatus consisting of a gray square measuring 40 × 40 × 40 cm (O'Hara & Co. Ltd., Tokyo, Japan) and illuminated with 100 lux at the center of the field. These tests were performed during the light period. Furthermore, a digital camera (CX5; Ricoh Company, Ltd., Tokyo, Japan) were used to

recode these tests. The movie recorded by the camera were utilized to measure the duration of events for each trait by human observation at a resolution of 0.1 s using tanaMove software version 0.01 (Goto et al. 2013). A single person who was different from the experimenter of the tame tests was watched and measured each phenotypic value using the recorded movie.

First I evaluated the level of motivation of an animal to approach human hands using the active tame test. In this test, each mouse was placed in the center of the field. The left hand of the experimenter, which was covered with a plastic glove, was placed at the bottom of the field and the fingers were continually moved slightly during the test to show the mouse that the hand was not just an inanimate object. Then, the experimenter kept the hand approximately 10 cm away from the tested mouse and followed the animal at the same distance when the mouse moved away. When the mouse headed toward the hand, the hand was kept stationary; thus, the mouse could approach the hand and even contact it. The durations of three behavioral traits: heading toward the hand, contacting the hand, and jumping, were measured. Locomotion, regarded as the duration of movement, including the heading and jumping behaviors described above, was also measured for each mouse.

The passive tame test to evaluate an animal's passive responses to a human hand was conducted for 1 min immediately after the active tame test. The experimenter again placed his hand in the test field and slowly chased the mouse. The experimenter attempted to touch the body of the mouse for as long as possible. The durations of three behavioral traits: heading toward the hand, accepting touching by the hand, and jumping, were measured. Locomotion was also measured following the same definition as used in the active tame test.

The stay-on-hand test was performed to evaluate the level of reluctance to avoid the human hand. This test was conducted for 1 min immediately after the passive tame test. The experimenter picked up the mouse by its tail using tweezers and put it on his hand. The thumb of the experimenter softly stroked the back of the mouse at a rate of once a second while the mouse stayed on the hand. The duration for which the mouse stayed on the hand was measured three times, the median of which was used to represent this trait.

1.2. 7. Association between behavioral indexes

To reveal the association between contacting and other behavioral indices, I used Spearman rank correlation coefficient due to the deviation from normal distribution in contacting (Shapiro-Wilk Normality Test, $P < 0.0001$). One hundred thirty seven mice from the C1 population at G_{12} were used in this analysis. Bonferroni correction was used for multiple comparisons for 8 times ($P < 0.006$). The analyses were performed using R (version 3.0.2).

1.3. Results

1.3.1. Phylogenic relationship between heterogeneous stocks

For establishing WHS, I chose eight wild strains that originated in different countries as founder strains (Table 1). To clarify genetic characteristics of the WHS, I performed genome-wide SNP analysis using the MegaMUGA array, which covers 77K SNPs (Welsh et al. 2012). The genotype data of the eight founder strains were compared with those of 45 other strains for which data are available at UNC Systems Genetics at

the University of North Carolina (http://csbio.unc.edu/CCstatus/index.py?run=GeneseekMM). The neighbor-joining (NJ) tree using 10,598 genome-wide SNPs revealed that the WHS founders have different genetic characteristics from other HSs and DO (Figure 1). The NJ tree of 53 inbred strains shows the genetic relationships for three subspecies groups: *domesticus*, *musculus*, and *castaneus*. Six strains (A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, and DBA/2J) were used as founders of both BHS and NHS and the remaining two founders of these stocks are derived from the same subspecies group, *domesticus*; thus, these two HSs are genetically very similar to each other. In contrast, the founder strains for WHS and DO originated from three different subspecies groups, *domesticus*, *castaneus*, and *musculus*, indicating that there is greater genetic diversity in WHS and DO than in other HSs (Figure 1).

Among the eight founder strains of WHS, five strains (BLG2, CHD, KJR, MSM, and NJL) are classified into the *musculus* subspecies group. This contrasts with DO in that, six strains (A/J, C57BL/6J, NOD/LtJ, NZO/HiLtJ, 129S1/SvImJ, and WSB/EiJ) out of eight are classified into the *domesticus* subspecies group in DO. This difference in founder strains suggests that the genetic characteristics of WHS differ from those of DO. These points indicate that WHS has high genetic diversity compared with BHS and NHS, and different genetic characteristics compared with the other stocks.

1.3.2. Selective breeding for active tameness

A novel outbred stock, WHS, was established by crossing the eight founder strains in line with the rotation rule, followed by random crossing (Figure 2). To clarify the genetic basis of active tameness, I conducted selective breeding for an index of contacting in the active tame test (Figure 3). The original population of WHS, which

was kept as 16 pairs, was doubled at the G_2 generation and split to form a second population. This was to avoid the loss of alleles due to random drift as the population size was relatively small. Then, selective breeding was initiated at the G_3 generation in one of these two populations, which was named S1, while the other population that was not selected was named C1, a control population (Figure 3). Another selected population was split from C1 and selection was initiated at the G_5 generation; this was named S2. In addition, another non-selection control population, C2, was split from S1 at the G_5 generation after selection for two generations and kept without any selection thereafter (Figure 3).

Figure 4 shows the transitions of nine behavioral indices including contacting, which is a selection index, through ten generations. Regarding contacting in the active tame test, the Steel–Dwass test, which is a non-parametric multiple comparison test, showed that the behavioral indices exhibit significant differences between C1 and S1, and between C1 and S2 after G_9 ($P < 0.001$; Figure 4; Table 2). In contacting, there were approximately three- and sixfold differences among the mean scores of S1 (3.44±0.35 (SEM)) and S2 (7.17±0.64), and the two control populations (1.14±0.14 and 0.99±0.13 for C1 and C2, respectively). One behavioral index, in the active tame test, "locomotion", did not show a significant difference (Table 2).

To reveal the correlation between contacting, an index of selection, and other behavioral indices, I used the Spearman's rank correlation coefficient. Since there is no sex difference within all four populations at G_{12} (Wilcoxon test, $P > 0.05$), no correction for sex was performed. For the analysis, I used 137 mice from the C1 population at G_{12} to avoid the selection effect on the analysis. After the Bonferroni correction ($P = 0.0062$), contacting showed associations with "locomotion" in active tame test ($P =$

0.0001, rho = 0.333) and passive tame test (P = 0.0033, rho = -0.250), “jumping” in passive tame test (P = 0.0008, rho = -0.284) and “staying” in the stay-on-hand test (P = 0.0036, rho = 0.247). Four other behavioral indices did not show any significant association with contacting (Table 3).

1.3.3. Haplotype contribution of heterogeneous stock

To characterize haplotype contribution in each strain for each region, I conducted genome-wide SNP analyses with the GigaMUGA array, which covers 144K SNPs in the mouse genome (Didion et al. 2014). In the SNP typing for the eight founder strains of WHS, one-strain-specific SNPs were mostly found in *domesticus* strains, PGN2 (n = 8,109 SNP) and BFM/2 (n = 6,038 SNP), followed by the *castaneus* strain HMI (n = 4,533 SNP). The five *musculus* strains had fewer one-strain-specific SNPs (Table 4). Given that the GigaMUGA array has been optimized for the analysis of *domesticus* mice, fewer SNPs for distinguishing *musculus* strains are included in the array. The genomic contribution of the founder strains to each population of WHS mice at the G_{12} generation was analyzed using the information on one-strain-specific SNPs. In 32 WHS mice from each of four populations, S1, S2, C1, and C2, at G_{12}, the genetic contributions of eight founder strains in the genome were inferred to be between 8.7% and 16.4% (Figure 5 and Figure 6-9 in detail). In the C1 population, in which no selective breeding had been applied, the contributions were inferred to be between 10.4% (HMI) and 16.0% (BLG2).

1.4. Discussion

1.4.1. Genetic characteristics of WHS founders and WHS

I conducted a phylogenetic analysis using SNP data to clarify the genetic structure of the founder strains of WHS and other mouse resources, as well as other inbred strains including the founder of other three outbred stocks. The topology of the phylogenetic tree constructed in the current study for inbred strains is similar to that reported in a previous study (Didion and de Villena 2013), suggesting that the phylogenetic relationship obtained from the MegaMUGA SNP data is reliable. The phylogenetic tree showed that the genetic characteristics of WHS founder strains differed from those of other HSs or DO founder strains.

In GigaMUGA SNP arrays, there were large differences in the numbers of one-strain-specific SNPs between *domesticus* and *castaneus* strains and *musculus* strains (Table 4). This may be due to the biased selection of polymorphic SNPs for distinguishing a variety of strains in *domesticus* subspecies group. By including more SNPs that enable *musculus* strains to be distinguished, SNP typing will allow us to conduct genetic analysis at higher resolution, especially in the case of using WHS.

Given that wild strains with different geographic origins showed higher genetic diversity (Koide et al. 2000), WHS made from eight wild strains is advantageous for identifying genes associated with complex traits. Haplotype inference for WHS mice revealed the genetic contribution of the founder strains in each region. The existence of genetic heterogeneity and random contributions from eight founder strains in the original WHS stock is one of the most important factors for identifying the genes affecting a trait by GWAS or selection mapping (Yalcin et al. 2010; Koide et al. 2012). In some regions, such as the 10-50 Mb of chromosome 6 in C1 population, the genetic

contributions of the eight strains were similar, although other regions, such as the 50-60 Mb of chromosome 18, showed different contributions among the strains (Figure 6). In the case of Collaborative Cross, a inbred mouse strains derived from outbred stock, the frequency of allele from one particular strain, WSB/EiJ, was significantly increased in the 56.1-Mb region on chromosome 2 in the three populations, suggesting a preference for the WSB/EiJ allele at this locus in Collaborative Cross (Collaborative Cross C 2012). However, analysis in the current study for the four populations, S1, S2, C1, and C2, revealed no significant increase in allele frequency for a particular strain in any of the populations. This suggested that there is little or no preference for a particular allele from any strain in mating of the WHS.

1.4.2. Selective breeding for active tameness

To understand the genetic basis of active tameness in further analysis, I conducted selective breeding of two WHS populations for contacting, which is a behavioral index of being motivated to approach humans (Goto et al. 2013). Notably, some other behavioral indices also changed through the selective breeding on contacting. Heading toward human hand in both active and passive tame tests was increased in the selected populations compared with that in the control populations. In addition, the duration of staying in the stay-on-hand test was also increased in the selected populations through the generations. Furthermore, I found significant correlation between contacting and locomotion in active tame and passive tame tests, jumping in passive tame test as well as staying in the stay-on-hand test. These findings implied that there are pleiotropic effects of genetic variation associated with contacting, or the same behavioral component was shared among contacting and those two behavioral indices. Since the

jumping behavior could be observed frequently in wild strains yet little or no in laboratory strain (Holmes et al. 2000; Fernandes et al. 2004; Goto et al. 2013), decreasing the duration of jumping in both selected populations suggested wildness of mouse should have decreased through the selective breeding. Further analyses of behaviors and neurochemical levels associated with these tame behaviors would potentially provide information that can help us to understand the neural and genetic bases of tameness in mice.

CHAPTER 2: Selective breeding mediated genetic mapping

2.1. Introduction

Previously, genetic analyses using genome-wide association studies (GWAS) and quantitative trait locus (QTL) mapping were conducted to identify the genetic bases of tameness in rats and foxes (Albert et al. 2009; Kukekova et al. 2011). In these analyses, however, candidate genes were not clearly identified and that could be caused by using small number of genetic markers and limited data accumulation. On the other hand, recently reported selection mapping studies, a method identifies loci using populations that have been subjected to natural/artificial selection with recombination, used numerous genetic markers. These studies have been conducted for domesticated and wild animals which revealed genetic loci potentially associated with tameness (Carneiro et al. 2014; Montague et al. 2014). However, little evidence for the association between the candidate loci and the actual behavioral phenotype has been reported. This is because these studies mainly focused on genetic polymorphisms and conducted few experiments at the behavioral traits. In addition, these methods were affected by the genetic structures due to uncontrolled migration of the compared populations. Taken together, identifying genetic basis of tameness requires more refined approach.

In contrast to existing domesticated animal populations, mouse as a model organism can be a useful resource for selection mapping to identify genetic basses of tameness. Several advantages of mice include easily control of genetic and environmental heterogeneity affecting genetic mapping analysis, and accessibility of genomic data and large amount of data on the current databases.

Here I conducted selection mapping using the established WHS in order to identify genetic loci associated with tameness. Because the genome of WHS contains heterogeneity but the founder and mating scheme is already-known, I applied these information to the computer simulation for selection mapping. Following selection mapping, I performed association mapping to confirm the association between the selected haplotype and selected index, contacting. Then I also search candidate genes within the region potentially affects contacting.

2.2. Methods

2.2.1. Computer simulation for identification of selected regions

I hypothesized that the allele frequency for particular SNPs associated with tameness should increase through the generations by the selective breeding. Due to the usage of large number of SNPs, the multiple testing problem should be addressed. The threshold becomes smaller and the statistical power falls small as the number of SNPs increases while adjusting the standard methods, such as using the Bonferroni correction. Additionally it is required to handle the linkage disequilibrium for SNPs linked on the same chromosome witch are transmitted together. To resolve the problems, I used a statistical test to increase allele frequency of SNP by whole-genome simulation. Here I simulated the frequency of an SNP within a population in a particular generation. Strain specific SNPs were subjected to the simulation, and a statistical test using a no-selection model was used to detect an increase in allele frequency. The simulation involved three setting: (1) strain specific 19 autosomal SNP, determined by GigaMUGA, were used. (2) Actual pedigree in the mating in the experiment was used for the simulation. (3)

Physical position of each SNPs were converted to the positions in the genome according to their genetic distances obtained from the Mouse diversity array (Yang et al. 2009).

I run a simulation following assumptions while the above settings used: (1) recombination occurred in the genome at random. (2) After recombination, the 19 autosomal chromosomes were independently inherited by the offspring according to the actual mating scheme. (3) No artificial/natural selection acted on the SNPs. (4) the maximum frequency of all SNPs was recorded at G_{12}. I repeated the process with 10,000 times. The 10,000 recorded frequencies from the simulation were used as the null distribution to detect the selection. The genome-wide significance level was set at the 5^{th} percentile of the null distribution. The process was performed for four populations (C1, C2, S1, and S2).

2.2.2. Association analysis

When I find the selected SNP and the allele of the SNP should be associated with contacting, which is a major selection index. Therefore I confirmed association between the allele in the selected SNP and contacting value.

For the association analysis, I used different model in each population. This is because the phenotypic distribution of contacting was differed among populations and the pattern of the distribution could affect for the result of each test. Here I used two models, one was mixed linear model witch assumed the standard Normal distribution, and the other was mixed Cox model which is corrected for the value includes many 0. Family data was used as mixed effects. The multiple comparison problem was corrected by using a false discovery rate ($q = 0.10$). Due to the linkage disequilibrium, I considered any region near the SNPs that exceed the threshold as a candidate region for

tameness. R v. 3 and additional R packages, lme4 v1.1.7 and qvalue v. 1.35.0. were used to perform association analysis.

2.2.3. Search for candidate genes

I searched the Mouse Genome Informatics database (MGI 6.02; accessed on February 3, 2016) for the detected region where exceeded the genome-wide significance level to identify the candidate genes. Focusing on contacting which is associated with active tameness, the genes within the detected region should be expressed in mouse brains. In addition the detected region should be associated with tame-related behavior, such as anxiety. Therefore I used keywords search in MGI for detected region using the terms including five terms (anxiety, fear, exploration, novelty, and social) as tame-related behaviors. For detecting the expression genes, I searched genes expressed in the brain for the ATR1 region using the Gene Expression Database (GXD) in MGI.

2.3. Results

2.3.1. Selection mapping of genetic loci for contacting

In the selective breeding for contacting, a significant increase of the behavioral score for contacting was observed at G_{12} (Figure 4a; Table 2). Especially, contacting of S2 population was higher than S1 population. To identify the genetic region associated with increased contacting, I conducted selection mapping using the SNP data from two selected populations, S1 and S2, and two control populations as controls (Figure 10).

In the mapping, I examined whether any locus showed a significant increase of a particular allele frequency that was greater than would be expected from random genetic

drift. I calculated the threshold using a computer simulation of allele frequencies in each generation based on a non-selection model using the pedigree and the SNP data for the eight founder strains. The thresholds for a significant increase of allele frequency were determined to be 0.703–0.797 (Table 5). In the selected population S2, the observed allele frequency of an SNP (UNC20197962) on chromosome 11 was 0.75 (Maximum allele frequency: $P = 0.0017$), which exceeded the threshold for the MSM strain (Table 5). In addition, heterozygosity decreased around the region where the selected SNP was present, suggesting that the selective sweep of this region occurred in the broad genomic region (Figure 11a-b). The other selected population, S1, and the control populations C1 and C2, had no SNPs for which the frequency significantly increased (Figure 10; Table 5). I found that the allele frequency of the SNP (UNC20197962) derived from MSM was zero in the S1 and C2 populations, and 0.188 in C1. These data suggest that the MSM allele of the SNP UNC20197962 had been lost due to genetic drift during the early stage of breeding before C2 was made from S1.

From the analysis using allele frequency for simple SNP genotype, no information on the length of the haplotype block was obtained. To characterize the haplotype block that increased in frequency during selective breeding, I inferred the haplotype using random forest (Maples et al. 2013). The major contribution to the inferred haplotype in this region is from the MSM strain (Figure 11b). The haplotype block of MSM including the detected SNP is approximately 52 Mb, between 67.7 and 119.7 Mb of chromosome 11.

2.3.2. Association mapping

I assumed that the MSM haplotype of the region should be associated with contacting. Applying the association analysis using mixed linear model for S2 and mixed cox model for C1, I found two smaller regions within the detected region, and named Active tameness regions, ATR1 (69,042,047bp - 81,256,559bp) and ATR2 (96,888,139bp - 104,407,876bp) (Figure 11c).

Tameness is potentially associated with levels of anxiety and fear toward humans or the human-mediated environment (Price 2002; Albert et al. 2008). Therefore, genes related to behaviors such as anxiety, fear, exploration, novelty, and social behavior could be associated with tameness. I searched genes for these five key words in the Mouse Genomics Informatics database and extracted 27 genes, which potentially affect tameness (Table 6).

2.4. Discussion

I identified the selected region on chromosome 11 based on allele frequencies that exceeded the threshold determined by the simulation for allele frequencies using pedigree and founder genotype data. Recent genomic studies in which selection mapping was conducted for domesticated (Akey et al. 2010; Wang et al. 2013) and wild animals (Poelstra et al. 2014) revealed genetic loci associated with animal domestication. Compared with the domesticated and wild animals, WHS is more useful to conduct selection mapping for complex traits. The first reason for this is that there are the advantages on easier availability of using pedigree information and genomic data on founder strains in the case of WHS, while these data are not available for wild animals

and most domesticated ones. The availability of the pedigree and genomic data for the founder strains allows us to perform more accurate fitting for the theory of laws of inheritance. The second reason is that the environmental conditions of the mice are controlled better for WHS than for other wild and domesticated animals. Given that environmental variance affects the phenotypic variance of the traits (Anholt and Mackay 2009), the circumstances in the laboratory, with the animals being kept in a highly similar environment, allowed us to minimize environmental effects. For this reason, the proportion of genetic variance relative to the phenotypic variance of the trait in the laboratory animals should be larger than that in non-laboratory populations. Based on the highly controlled environmental conditions for laboratory mice, selection mapping using WHS is a powerful method to reveal the genetic regions associated with complex traits.

In the results of selection mapping, large deviations in the frequency of each SNP genotype as well as the haplotype data were detected on chromosome 11. The combination of simulation of SNP allele frequencies and the actual frequencies of each SNP allele allowed us to find strong selective sweeps in the S2 population genome. This is because the region over the selected threshold detected by the simulation must be a locus associated with selection for tameness. The strong selection that occurred on the MSM allele on the region indicates that its effect on contacting might be large. In addition, the S1 population, which was selected for contacting and exhibited a significantly higher contacting score than the two control populations, C1 and C2, lacks the MSM allele of the ATR1 SNP (UNC20197962). In G_{12}, the contacting of S2 was significantly higher than that of the S1 population, suggesting that maintenance of the MSM allele of ATR1 increases contacting, although higher contacting can also be

induced by the combination of other loci without the MSM allele of ATRs. In the present study, I revealed that the MSM allele of ATRs has an effect on increasing contacting. In contrast, Goto et al. (2013) analyzed and compared contacting scores using a series of laboratory and wild strains, they showed that wild strain MSM exhibited a lower level of contacting than other wild strains. Therefore, the MSM allele of ATRs should interact with other genetic loci to produce a higher level of contacting in WHS.

Given that a large reduction of heterozygosity indicates the existence of selective pressure on the genome over the course of selective breeding, the two sub-regions in which heterozygosity was reduced implied that there are two different causative elements within ATRs that affect tameness. This could be one of the reasons why the detected region by the simulation exhibits a large haplotype block showing skewed allele frequencies.

ATRs includes 27 genes associated with tameness-related behavior (Table 6). One of the candidate gene located in ATR1 is solute carrier family 6, member 4 (*Slc6a4*), an integral membrane protein that transports the neurotransmitter serotonin and is involved in the pharmacological targeting of psychomotor stimulants such as amphetamine and cocaine. The gene was detected by 6 keywords by the searching. It has been reported that *Slc6a4* is associated with aggressive behavior (Holmes et al. 2003), social behavior (Page et al. 2009), and anxiety-related responses in mice (Narboux-Neme et al. 2011). Notably, in a previous study, selection mapping using genomic data from dogs and wolves showed that *Slc6a4* has been selected during the course of dog domestication (Wang et al. 2013). In the same paper, the association of variations of *SLC6A4* in humans with various neuropsychiatric symptoms, such as obsessive compulsive

disorder, depression, and autism, was reported. In addition, the downstream metabolite of a transport substrate of *Slc6a4* was also found to be associated with aggressive behavior in dogs in another association study (Reisner et al. 1996). Therefore, it is possible that *Slc6a4* also plays an important role in changing traits of active tameness during selective breeding in mice.

CHAPTER 3: Comparative genomic analysis with rats and dogs

3.1. Introduction

Tameness could be observed in various domesticated animals (Wright 2015), and the tameness might be a basic factor associated with domestication (Grandin and Deesing 2014), especially in the early stage of domestication (Wilkins et al 2014). Observing tameness in many animal species suggesting that the genetic component underliing tameness should be shared between the animal species, and the number of genetic factors could be small (Grandin and Deesing 2014). Revealing master genetic factor underling tameness contribute to understand how animals were domesticated, especially in the initial stage of domestication.

In Chapter 2, I found ATRs associate with tameness. If the genetic basis of tameness is shared between the domestic animal species, orthologs of causative genes in ATRs might also be associated with tameness and/or domestication in other animal species. Examples of the candidates for this kind of approach is regions associated with tameness in rats reported by QTL mapping (Albert et al. 2009, Heyne et al. 2014). Additionally, if the regions associated with tameness selected during the animal domestication, a signature of selection for the regions should be also observed in the case of dogs (Wang et al. 2013). Both studies provided many regions potentially associated with tameness in the animal species.

In the case of rats, Belyaev and his collaborators have selected tame and aggressive rats from 233 wild-caught animals to uncover genetic basses of tameness (Albert et al. 2008). Although the selection experiment is currently stopped due to low fertility (Albert et al. 2012) and no genome-wide selection mapping was performed for

them, Heyne et al. (2014) revealed genes altered expression might be influencing tameness. The study suggested that eight loci for tameness QTL in rats and hundreds of genomic loci that influence gene expression levels between tame and aggressive rats.

In another case, dog is one of the most intensely domesticated animals and exhibits a high level of tameness (Belyaev 1979). Tameness of dog has been selected in the course of domestication (Belyaev 1979) and then the tame related region should have been selected during the domestication. Several genomic analyses have revealed selected regions which have been selected during the domestication (Akey et al. 2010; Wang et al. 2013). Especially they have performed genomic comparisons between gray wolves and Chinese indigenous dogs (Wang et al. 2013), and among 10 different dog breeds (Akey et al. 2010). In addition, recently Ilska et al. (2017) carried out GWAS for 12 personality traits including tame-related traits using 885 Labrador Retrievers to identify the genetic basses of the traits. These results could help to identify the region potentially influence tameness in dogs.

In order to identify the shared genomic regions associated with tameness among three mammalian species, mice, rats, and dogs, I compared the ATRs in mice with homologous regions of rats and dogs where potentially associated with tameness which reported by the previous study.

3.2. Methods

I established simulation method to estimate the probabilities that ATR homologous regions in mice, and tameness QTL in rats and selected regions in dogs. In the simulation, randomized regions in the mouse genome (ATR1: 69,042,047bp -

81,256,559bp (12.21 Mb) and ATR2 (96,888,139bp – 104,407,876bp (7.51 Mb)) were collected and the gene data set was first prepared. The gene dataset was manually assembled from a total of 8 tameness QTL in rats (Heyne et al. 2014), 4 SNPs which suggestive level association by dog GWAS (Ilska et al. 2017) and 354 previously described segments as selected regions (Akey et al. 2010; Wang et al. 2013).

The information of homologous genes in mice, rats and dogs were obtained from Archive EnsEMBL release 67. The homologous positions in rat and dog genes was corresponding to the extracted mouse genes in the ATRs. The number of QTL and selected regions overlapped with the ATRs homologous region was then determined using the orthologous gene dataset. I actually found the one overlap for rats and three overlap for dogs. For rat analysis, the probabilities of observing were more than one overlap with the same length as ATR1 (12.21 Mb) and ATR2 (7.51 Mb), respectively. For dog analysis, because two and one overlaps were found in ATR1 and ATR2 respectively, the probabilities of observing were more than two or one overlaps with regions of the same lengths as ATR1 (12.21 Mb) and ATR2 (7.51 Mb), respectively. These simulation were performed 100,000 times and then the probabilities were estimated.

3.3. Results and discussion

I compared the ATRs in the mouse genome with homologous regions in the rat tameness QTL, dog tameness related regions and selected regions during the domestication. By using a total of 8 tameness QTLs in rats (Heyne et al. 2014), 4 SNP positions which have been identified as exceeding suggestive level by the dog GWAS

for tame related traits (Ilska et al. 2017) and 354 previously described segments as selected regions (Akey et al. 2010; Wang et al. 2013), overlap and its probabilities were estimated by the simulation based method (see methods).

As a result, I found four overlaps, one syntenic region for rat tameness QTL and three syntenic regions for selected regions in the dog (Table 7), and no overlap for dog GWAS results was identified. Although overlapping probability of rat tameness QTL was relatively high (ATR2: P = 0.214), the other two probabilities for dog selected regions were unlikely by chance (ATR1: P = 0.051, ATR2: P = 0 .096) (Table 7). The overlap between the mouse ATR2, syntenic regions of rat tameness QTL and selected region in dogs was detected an approximately 0.61Mb region on mouse Chromosome 11 (100,887,204bp – 101,500,066bp), named Shared Tameness Region, STR. These results suggested that the region might associate with tameness in these three species although no direct evidence supports this hypothesis at phenotypic level in dogs.

One of the reason why I could not identify any overlap to the GWAS regions in dogs associate with the tame related traits is a difficulty to identify the genetic region influencing the tameness related behavioral traits in dogs due to its complexity. The dog genomic study by Ilska et al. (2017) suggested that personality traits in dogs like are mainly influenced by many genetic regions with small effect size to the phenotype, so large datasets should be required to improve statistical power to identify the individual genes influencing the traits. They actually identified only small number of regions with "suggestive" level association, not significant level. Therefore, they might have missed to identify most of GWAS positive regions. More comprehensive dog GWAS with larger sample size might identify evidence of overlap with mouse tameness in the detected region.

Totally 25 protein coding and exist homologous genes among three species were located within STR. Two genes, *Naglu* and *Hsd17b1*, were remain after the extraction by the keyword search as shown in Table 6 (see Ch.2). Although expression of these genes in the brain have not been confirmed so far, *Naglu* knockout mouse decreased fear-related response (Li et al. 2000) and *Hsd17b1* knockout mouse exhibits decreased exploration in new environment (International Mouse Phenotyping Consortium (IMPC) website as 21 Nov 2017). Although no MSM specific variation which could affect amino acid sequence and expression within the genes was identified so far, both genes could be leading candidate genes and tame test will be required for evaluating whether the gene influence tameness or not.

In summary, comparing potentially influencing tameness in three species, mice, rats and dogs, I found an overlap region, STR, between three species. Further analyses will be needed to conduct molecular genetic analysis such as expression analysis for the genes within the STR to reveal the molecular genetic basses associated with tameness in mammalian species.

Acknowledgment

I would like to thank my superviser, Dr. Tsuyoshi Koide, for his appropriate advice, kindly support and other general things to spend my PhD course. I appreciate my progress committee members, Dr. Ituro Inoue, Dr. Naruya Saitou, Dr. Yasukazu Nakamura, Dr. Jun Kitano, and Dr. Naoki Osada for kind advice. I also thank Dr. Tatsuhiko Goto for the lecture about behavioral tests and providing the partial data of phenotype and genotype of WHS. Dr. Aki Takahashi and Dr. Kazuto Yoshimi gave me many advices through substantial discussions. Dr. Akira Tanave kindly helped me how to conduct the programing of computer and molecular experiments, and frequently discuss about the research. Dr. Hirofumi Nakaoka and Dr. Jo Nishino gave me useful advice and kindly support on genetic analyses using statistical model. Dr. Richard Mott and Dr. Jonathan Flint taught me how to analyze using genomic data in heterogeneous stock mouse. I appreciated all members of the animal facility at NIG for caring, Ms. Motoko Nihei for quantification of behaviors, and Ms. Naoki Takahashi and Ms. Akiko Tsuchiya for their kindly care of the mice.

References

Agnvall, B., Jöngren, M., Strandberg, E., Jensen, P. 2012. Heritability and genetic correlations of fear-related behaviour in Red Junglefowl–possible implications for early domestication. *PloS one*, 7: e35162.

Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, Nicholas TJ, Neff MW. 2010. Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci U S A* **107**: 1160-1165.

Albert FW, Shchepina O, Winter C, Rompler H, Teupser D, Palme R, Ceglarek U, Kratzsch J, Sohr R, Trut LN et al. 2008. Phenotypic differences in behavior, physiology and neurochemistry between rats selected for tameness and for defensive aggression towards humans. *Horm Behav* **53**: 413-421.

Albert FW, Carlborg O, Plyusnina I, Besnier F, Hedwig D, Lautenschlager S, Lorenz D, McIntosh J, Neumann C, Richter H et al. 2009. Genetic architecture of tameness in a rat model of animal domestication. *Genetics* **182**: 541-554.

Albert, F. W., Somel, M., Carneiro, M., Aximu-Petri, A., Halbwax, M., Thalmann, O., Blanco-Aguiar, JA., Plyusnina, I. Z., Trut, L., Villafuerte, R., Ferrand, N., Kaiser, S., Jensen, P., Pääbo. S. 2012. A comparison of brain gene expression levels in domesticated and wild animals. *PLoS genetics* 8(9), e1002962.

Anholt RR, Mackay TF. 2009. *Principles of behavioral genetics*. Academic Press.

Belknap JK, Haltli NR, Goebel DM, Lame M. 1983. Selective breeding for high and low levels of opiate-induced analgesia in mice. *Behav Genet* **13**: 383-396.

Belyaev DK. 1979. Destabilizing selection as a factor in domestication. *Journal of Heredity* **70**: 301-308.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and

missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**: 1084-1097.

Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alfoldi J, Barrio AM, Pielberg G, Rafati N, Sayyab S, Turner-Maier J et al. 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**: 1074-1079.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**: 7.

Chia R, Achilli F, Festing MF, Fisher EM. 2005. The origins and uses of mouse outbred stocks. *Nat Genet* **37**: 1181-1186.

Collaborative Cross C. 2012. The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics* **190**: 389-401.

DeFries JC, Gervais MC, Thomas EA. 1978. Response to 30 generations of selection for open-field activity in laboratory mice. *Behav Genet* **8**: 3-13.

Didion JP, Buus RJ, Naghashfar Z, Threadgill DW, Morse HC, 3rd, de Villena FP. 2014. SNP array profiling of mouse cell lines identifies their strains of origin and reveals cross-contamination and widespread aneuploidy. *BMC genomics* **15**: 847.

Didion JP, de Villena FP. 2013. Deconstructing Mus gemischus: advances in understanding ancestry, structure, and variation in the genome of the laboratory mouse. *Mamm Genome* **24**: 1-20.

Fernandes C, Liu L, Paya-Cano JL, Gregorova S, Forejt J, Schalkwyk LC. 2004. Behavioral characterization of wild derived male mice (Mus musculus

musculus) of the PWD/Ph inbred strain: high exploration compared to C57BL/6J. *Behav Genet* **34**: 621-630.

Flint J, Eskin E. 2012. Genome-wide association studies in mice. *Nat Rev Genet* **13**: 807-817.

Goto T, Tanave A, Moriwaki K, Shiroishi T, Koide T. 2013. Selection for reluctance to avoid humans during the domestication of mice. *Genes Brain Behav* **12**: 760-770.

Grandin T, Deesing MJ. 2014. Behavioral genetics and animal science. In Grandin T, Deesing MJ. (Eds.). (2013). Genetics and the behavior of domestic animals. Academic press. UK.

Grahame NJ, Li, T. K., Lumeng, L. 1999. Selective breeding for high and low alcohol preference in mice. *Behavior genetics* **29**: 47-57.

Heyne, H. O., Lautenschläger, S., Nelson, R., Besnier, F., Rotival, M., Cagan, A., Kozhemyakina, R., Plyusnina, I. Z., Trut, L., Carlborg, Ö., Petretto, E., Kruglyak, L., Pääbo, S., Schöneberg, T., Albert, F. W. 2014. Genetic influences on brain gene expression in rats selected for tameness and aggression. *Genetics* 198: 1277-1290.

Hitzemann R, Edmunds S, Wu W, Malmanger B, Walter N, Belknap J, Darakjian P, McWeeney S. 2009. Detection of reciprocal quantitative trait loci for acute ethanol withdrawal and ethanol consumption in heterogeneous stock mice. *Psychopharmacology (Berl)* **203**: 713-722.

Holmes A, Parmigiani S, Ferrari PF, Palanza P, Rodgers RJ. 2000. Behavioral profile of wild mice in the elevated plus-maze test for anxiety. *Physiology & behavior* **71**: 509-516.

Holmes A, Murphy DL, Crawley JN. 2003. Abnormal behavioral phenotypes of serotonin transporter knockout mice: parallels with human anxiety and depression. *Biological psychiatry* **54**: 953-959.

Ilska, J., Haskell, M. J., Blott, S. C., Sánchez-Molano, E., Polgar, Z., Lofgren, S. E., Clements, N. D. Wiener, P. 2017. Genetic Characterization of Dog Personality Traits. *Genetics* 206: 1101-1111.

Koide T, Goto T, Takano-Shimizu T. 2012. Genomic mixing to elucidate the genetic system of complex traits. *Exp Anim* **61**: 503-509.

Koide T, Moriwaki K, Ikeda K, Niki H, Shiroishi T. 2000. Multi-phenotype behavioral characterization of inbred strains derived from wild stocks of Mus musculus. *Mamm Genome* **11**: 664-670.

Kukekova AV, Trut LN, Chase K, Kharlamova AV, Johnson JL, Temnykh SV, Oskina IN, Gulevich RG, Vladimirova AV, Klebanov S et al. 2011. Mapping Loci for fox domestication: deconstruction/reconstruction of a behavioral phenotype. *Behav Genet* **41**: 593-606.

Li HH, Yu WH, Rozengurt N, Zhao HZ, Lyons KM, Anagnostaras S, Fanselow MS, Suzuki K, Vanier MT, Neufeld EF. 1999. Mouse model of Sanfilippo syndrome type B produced by targeted disruption of the gene encoding alpha-N-acetylglucosaminidase. *Proc Natl Acad Sci U S A*, 96: 14505-14510.

Lynch CB. 1980. Response to divergent selection for nesting behavior in Mus musculus. *Genetics* **96**: 757-765.

Maples BK, Gravel S, Kenny EE, Bustamante CD. 2013. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* **93**: 278-288.

McClearn, Gerald E., James R. Wilson, Meredith. W. 1970. The use of isogenic and heterogenic mouse stocks in behavioral research. In *Contributions to behavior-genetic analysis: The mouse as a prototype* (ed. G Lindzey, DD Thiessen), pp. 3-22. Appleton-Century-Crofts, New York,.

Montague MJ, Li G, Gandolfi B, Khan R, Aken BL, Searle SM, Minx P, Hillier LW, Koboldt DC, Davis BW et al. 2014. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci U S A* **111**: 17230-17235.

Narboux-Neme N, Sagne C, Doly S, Diaz SL, Martin CB, Angenard G, Martres MP, Giros B, Hamon M, Lanfumey L et al. 2011. Severe serotonin depletion after conditional deletion of the vesicular monoamine transporter 2 gene in serotonin neurons: neural and behavioral consequences. *Neuropsychopharmacology* **36**: 2538-2550.

Page DT, Kuti OJ, Prestia C, Sur M. 2009. Haploinsufficiency for Pten and Serotonin transporter cooperatively influences brain size and social behavior. *Proc Natl Acad Sci U S A* **106**: 1989-1994.

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Muller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**: 1410-1414.

Price EO. 2002. *Animal domestication and behavior*. CABI Pub., Wallingford, Oxon, UK ; New York.

Reisner IR, Mann JJ, Stanley M, Huang YY, Houpt KA. 1996. Comparison of cerebrospinal fluid monoamine metabolite levels in dominant-aggressive and non-aggressive dogs. *Brain Res* **714**: 57-64.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol biol evol* 4: 406-425.

Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, Palmer AA, McMillan L, Churchill GA. 2012. High-resolution genetic mapping using the Mouse Diversity outbred population. *Genetics* **190**: 437-447.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725-2729.

Threadgill DW, Hunter KW, Williams RW. 2002. Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. *Mamm Genome* **13**: 175-178.

Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS, Rawlins JN, Mott R, Flint J. 2006. Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet* **38**: 879-887.

Wang GD, Zhai W, Yang HC, Fan RX, Cao X, Zhong L, Wang L, Liu F, Wu H, Cheng LG et al. 2013. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun* **4**: 1860.

Welsh CE, Miller DR, Manly KF, Wang J, McMillan L, Morahan G, Mott R, Iraqi FA, Threadgill DW, de Villena FP. 2012. Status and access to the Collaborative Cross population. *Mamm Genome* **23**: 706-712.

Wilkins AS, Wrangham RW, Fitch WT. 2014. The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*, 197: 795-808.

Wright, D. 2015. The genetic architecture of domestication in animals. *Bioinformatics and biology insights* 9.Suppl 4: 11-20.

Yalcin B, Flint J. 2012. Association studies in outbred mice in a new era of full-genome sequencing. *Mamm Genome* **23**: 719-726.

Yalcin B, Nicod J, Bhomra A, Davidson S, Cleak J, Farinelli L, Osteras M, Whitley A, Yuan W, Gan X et al. 2010. Commercially available outbred mice for genome-wide association studies. *PLoS Genet* **6**: e1001085.

Yang H, Ding Y, Hutchins LN, Szatkiewicz J, Bell TA, Paigen BJ, Graber JH, de Villena FP, Churchill GA. 2009. A customized and versatile high-density genotyping array for the mouse. *Nature methods* **6**: 663-666.

Zombeck JA, Deyoung EK, Brzezinska WJ, Rhodes JS. 2011. Selective breeding for increased home cage physical activity in collaborative cross and Hsd:ICR mice. *Behav Genet* **41**: 571-58

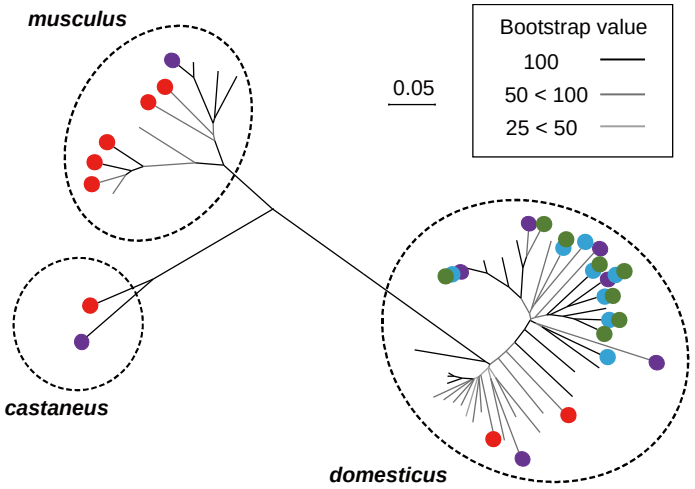

Figure 1. Neighbor-joining tree between eight WHS founder mouse strains and other inbred strains. Red, light blue, green, purple circles indicate founder strains of WHS, Boulder HS, Northport HS and Diversity Outbred, respectively. The bootstrap test was performed 1,000 times. A total of 10,598 SNPs obtained by MegaMUGA array were used. As the SNPs of the array were selected to show genetic differences among *M. m. domesticus* subspecies, the genetic difference among *M. m. musculus* subspecies seems smaller than the actual genetic distance. The tree was reconstructed using 53 strains (AEJ/GnLeJ, BFM2/Ms, A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, DBA/2J, I/LnJ, RIIIS/J, BLG2/Ms, 129S1/SvImJ, CAST/EiJ, NOD/ShiLtJ, NZO/HILtJ, PWK/PhJ, WSB/EiJ, CHD/Ms, CZECHI/EiJ, CZECHII/EiJ, DDY/Jcl, FVB/NJ, FVB, HMI, JE/LeJ, JF1/Ms, KJR/Ms, KK/HlJ, LEWES/EiJ, LG/J, LP/J, LT/SvEiJ, MOLF/EiJ, MRL/MpJ, MSM/Ms, CBA/CaJ, LP/J, NJL/Ms, NZW/LacJ, P/J, PERC/EiJ, PGN2/Ms, PWD, RBB/DnJ, RBF/DnJ, SF/CamEiJ, SH1/LeJ, SOD1/EiJ, ST/bJ, SWR/J, TIRANO/EiJ, WLA/Pas, WMP/Pas, ZALENDE/EiJ).

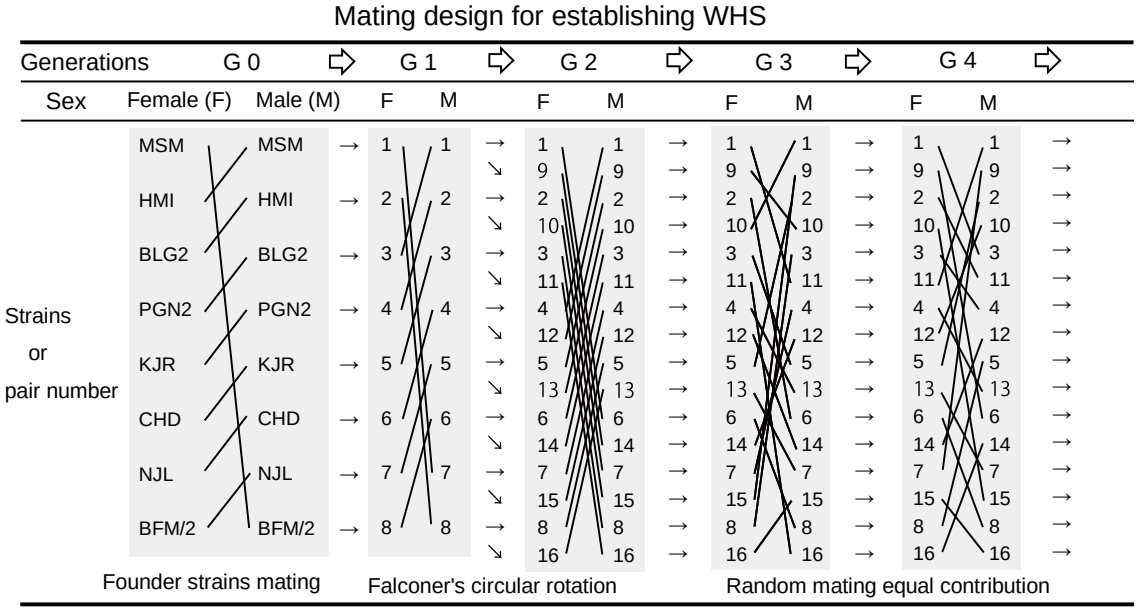

Figure 2. Mating design for establishing WHS. Mating scheme for the early stage of establishing WHS. Lines between founder strains or numbers indicate mating to produce progeny. The cage numbers in the next generations are given for the female lineage; thus, both male and female progeny were given the same cage number as their mothers, except for in G_1, in which the number was given for each female founder strain as shown in the figure.

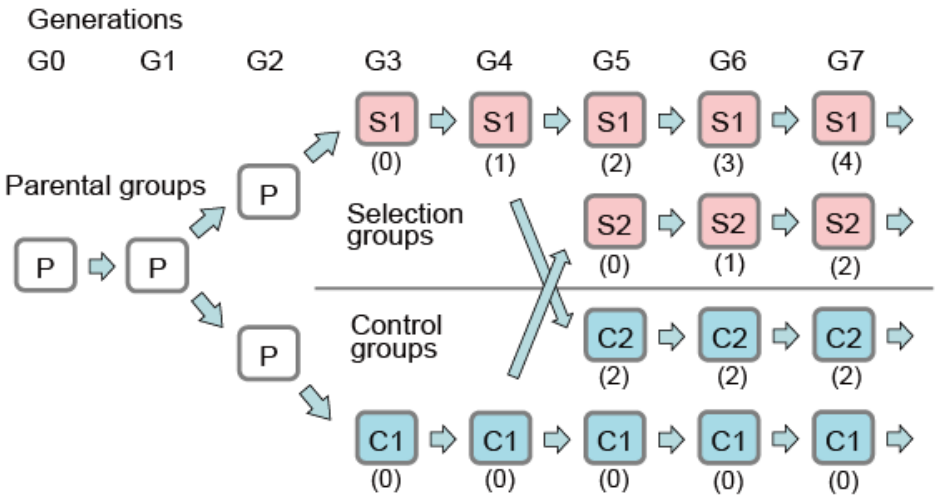

Figure 3. Mating design for selective breeding. b. The process for establishing two selection groups and two control groups. The numbers with parenthesis indicate the number of selection.

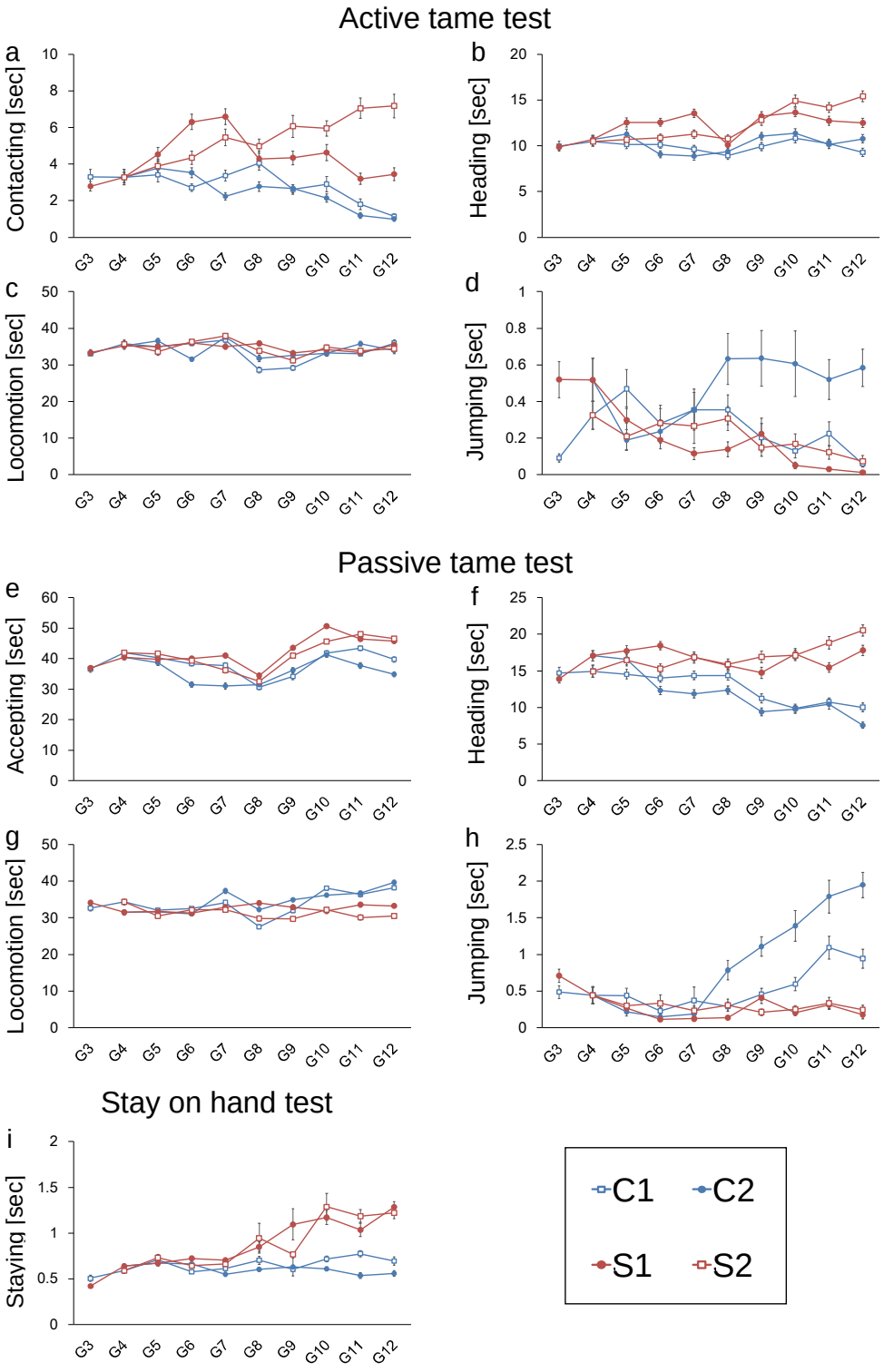

Figure 4. The fluctuation of each behavioral trait through the generations. Red lines and blue lines indicate selection and control groups, respectively. Error bars indicate SEM.

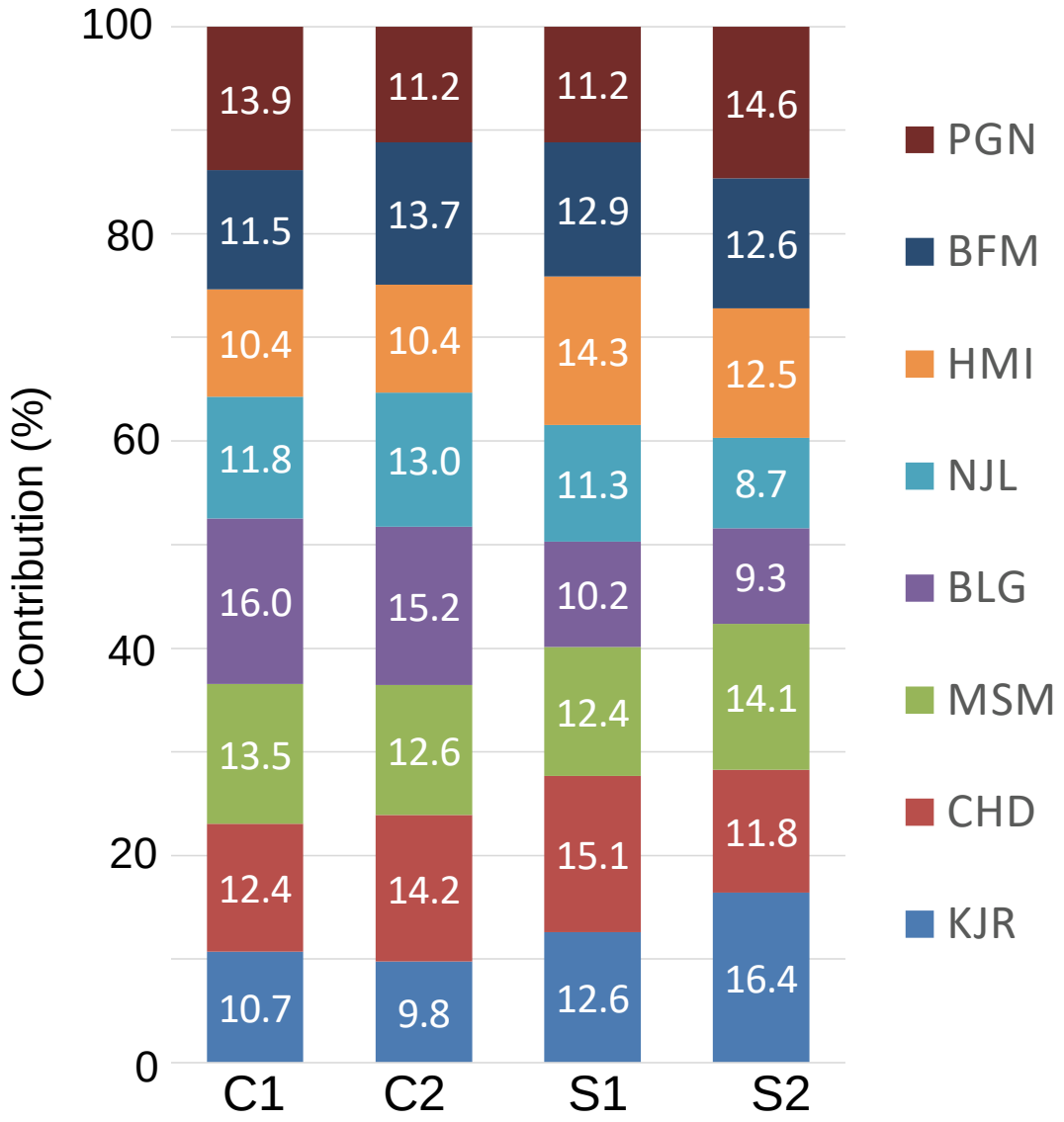

Figure 5. Genome-wide contribution of eight founder strains. The inferred haplotype for the eight founder strains using 60,321 autosomal SNPs was used. Numbers in each cell indicate the contribution of each founder strain.

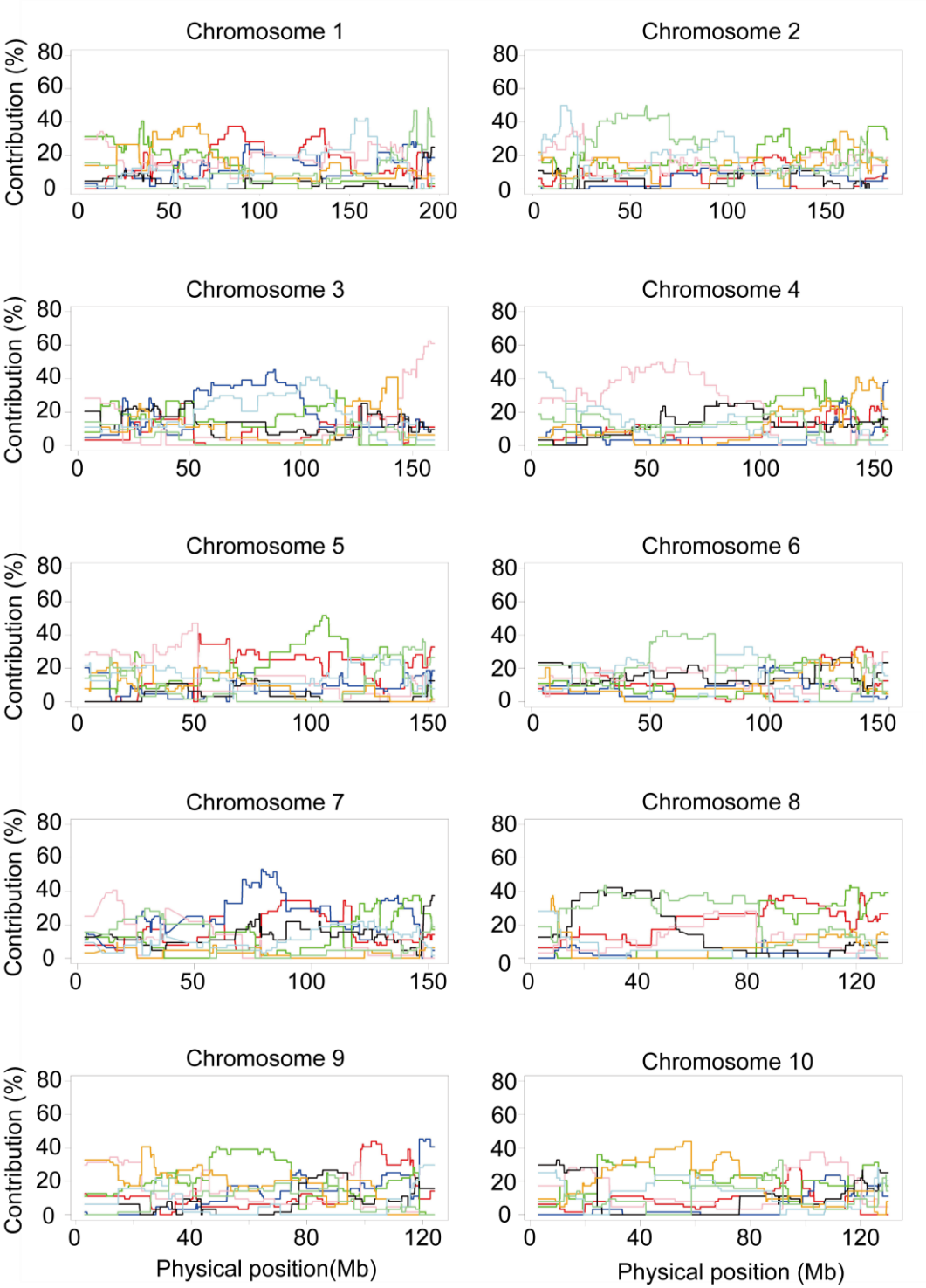

Figure 6. Contribution of each founder strain in the C1 population at G_{12}.

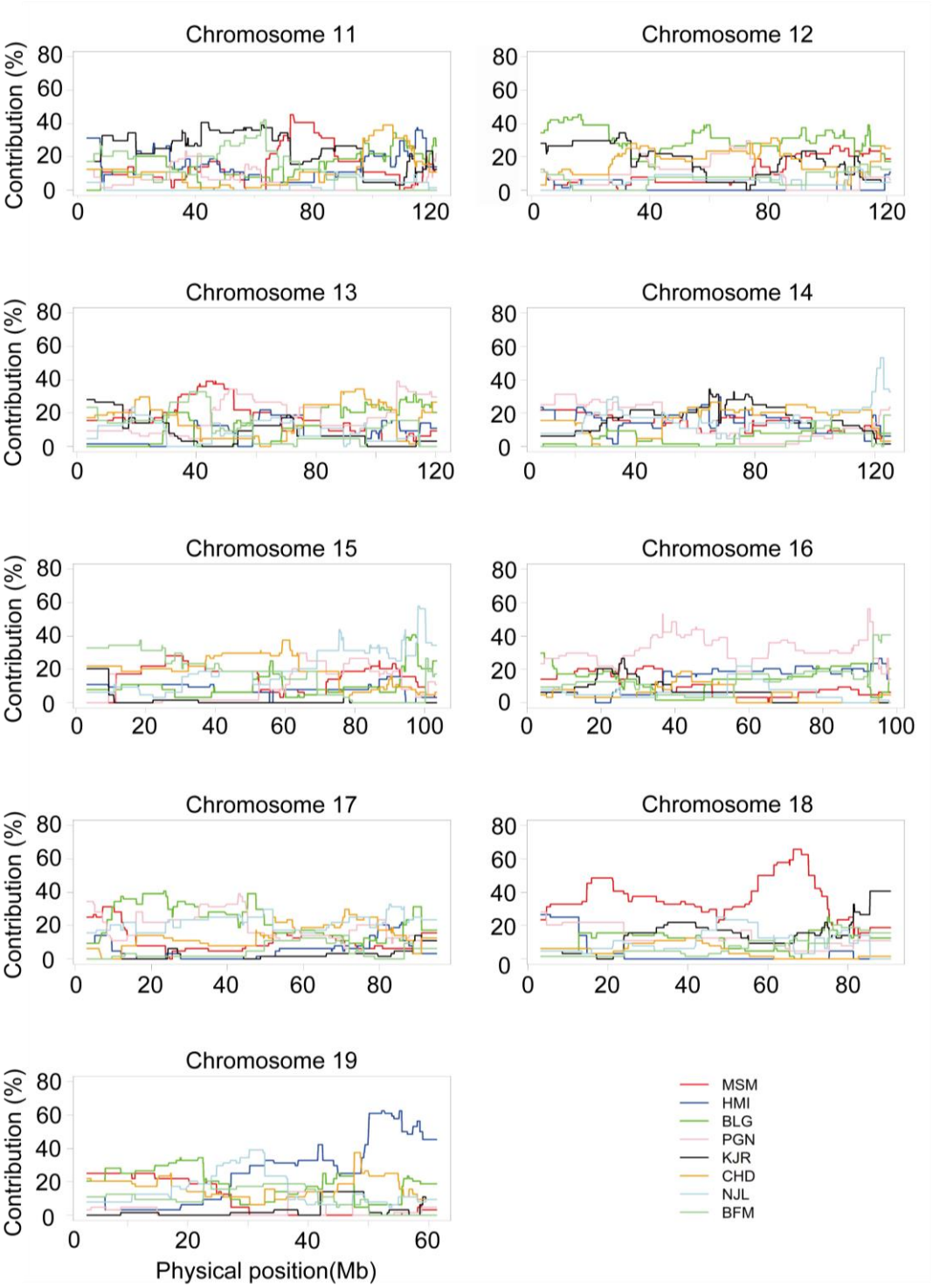

Figure 6.Continued.

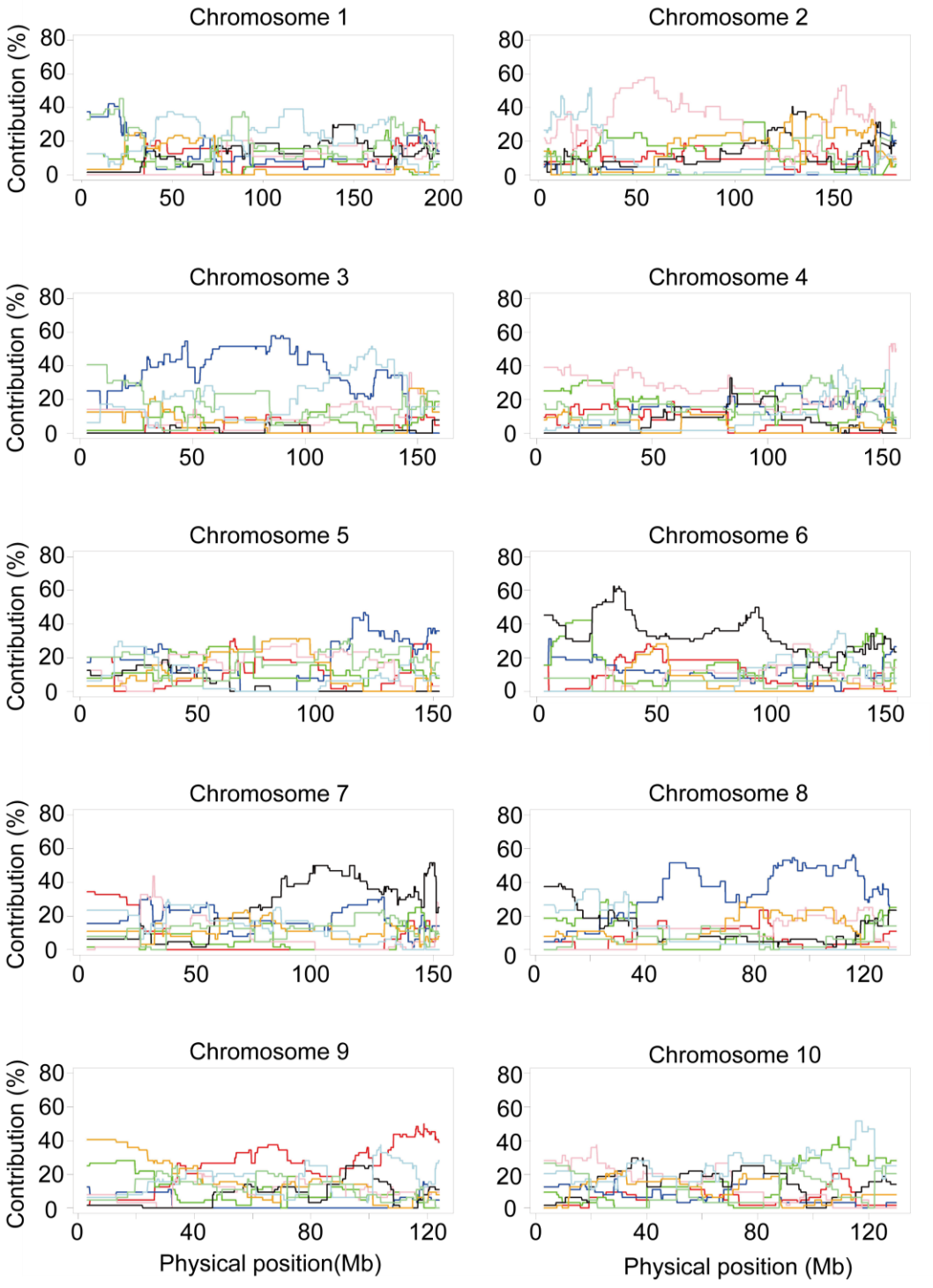

Figure 7. Contribution of each founder strain in the C2 population at G_{12}.

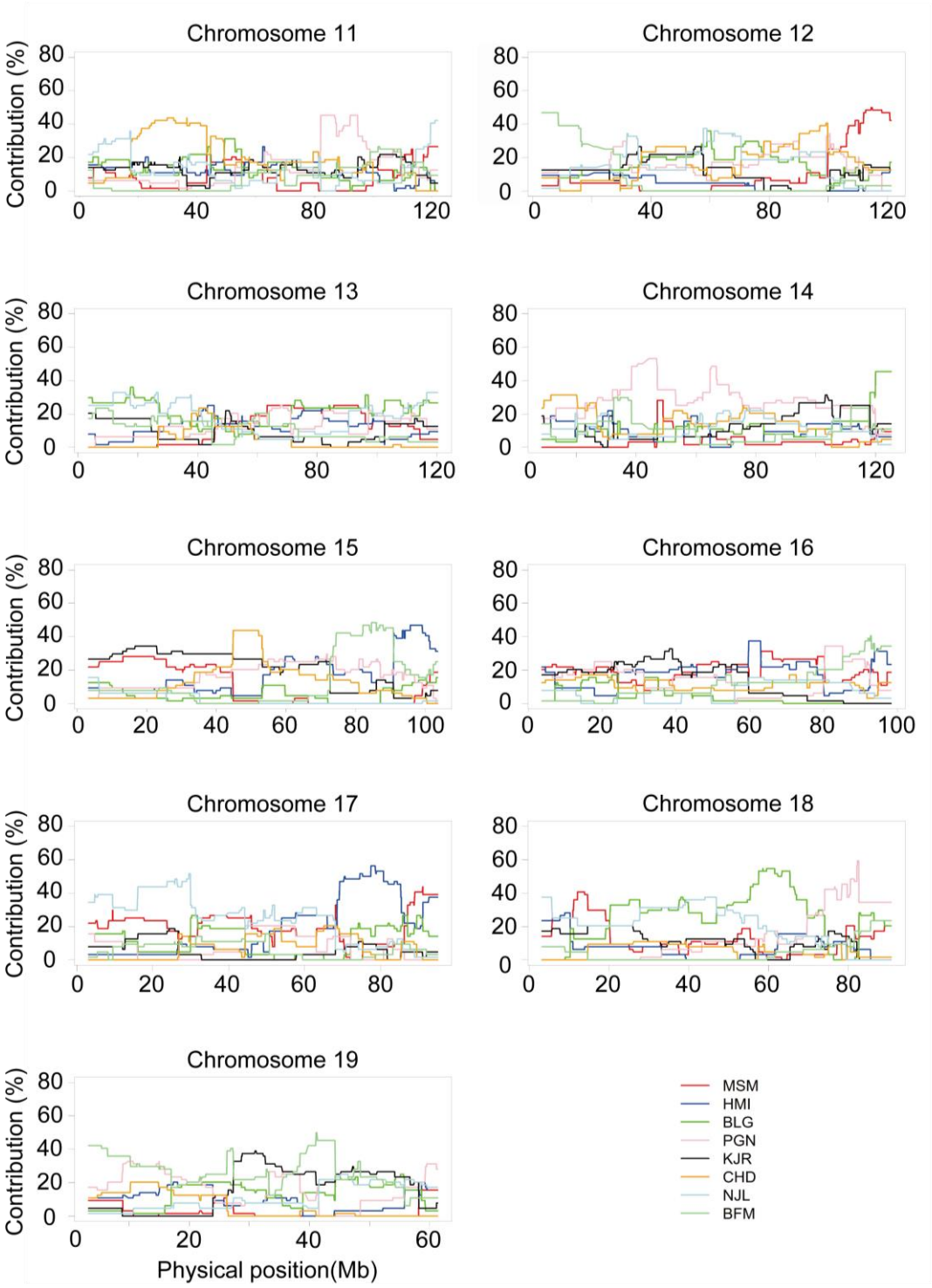

Figure 7. Continued.

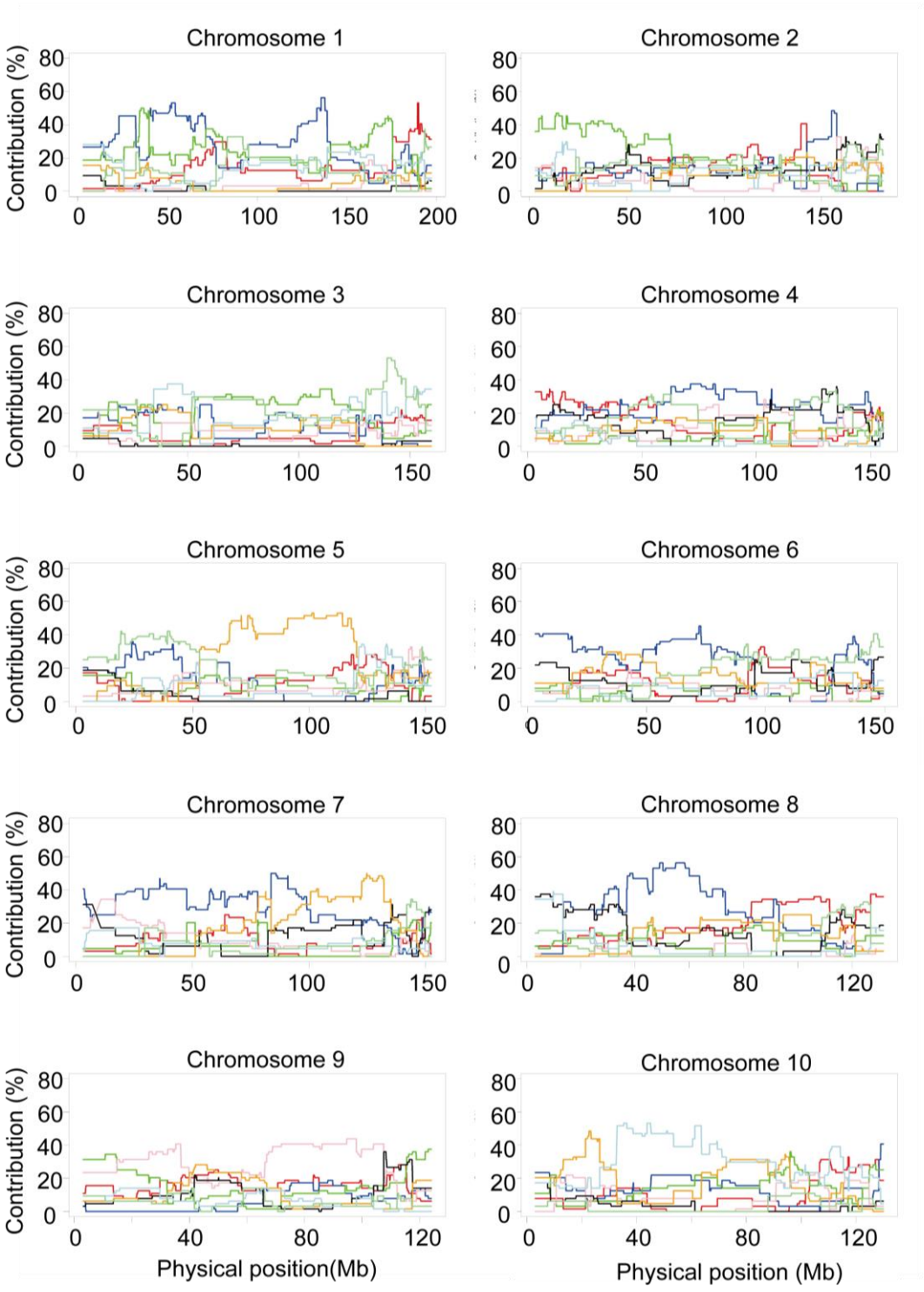

Figure 8. Contribution of each founder strain in the S1 population at G_{12}.

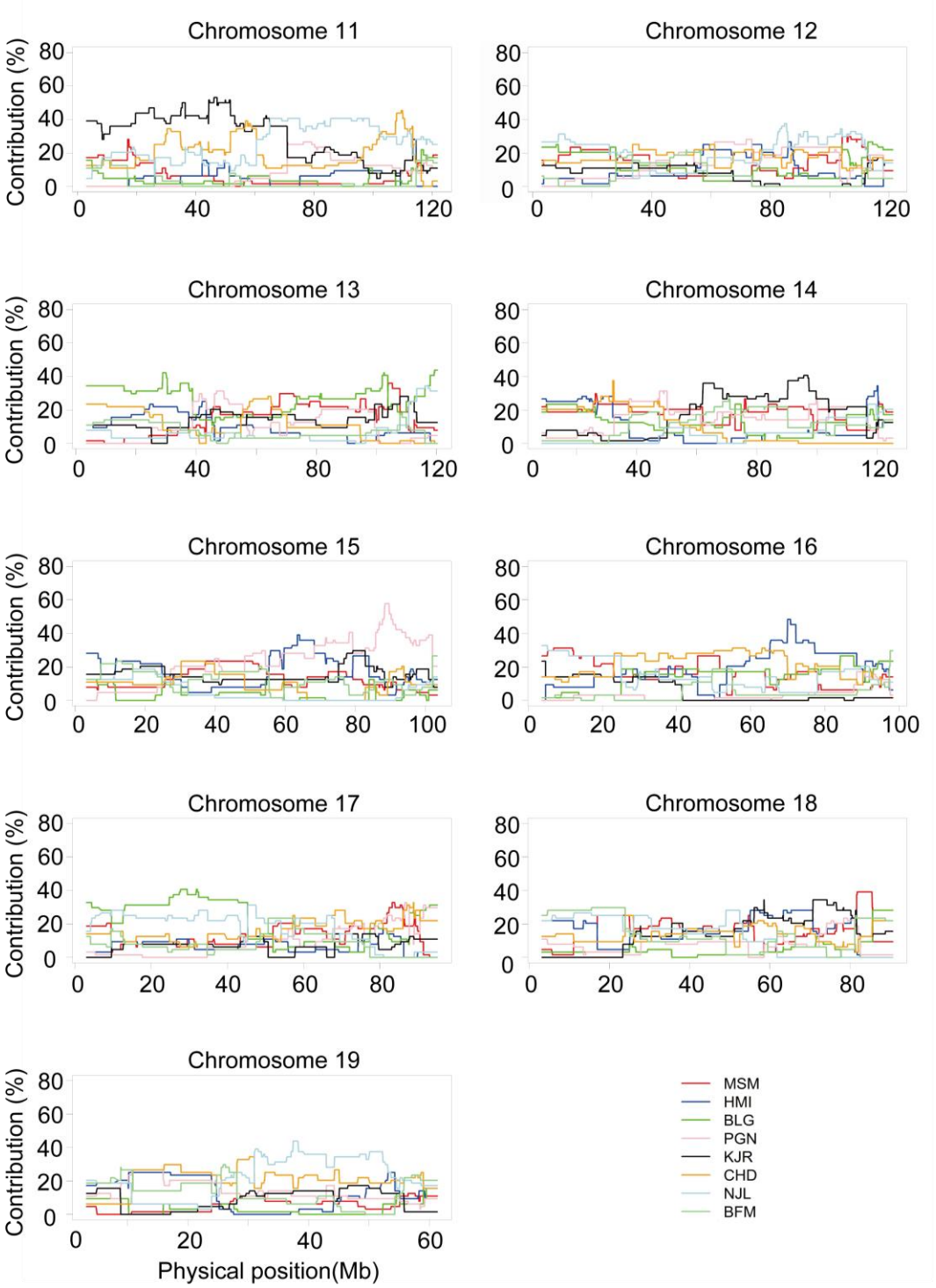

Figure 8. Continued.

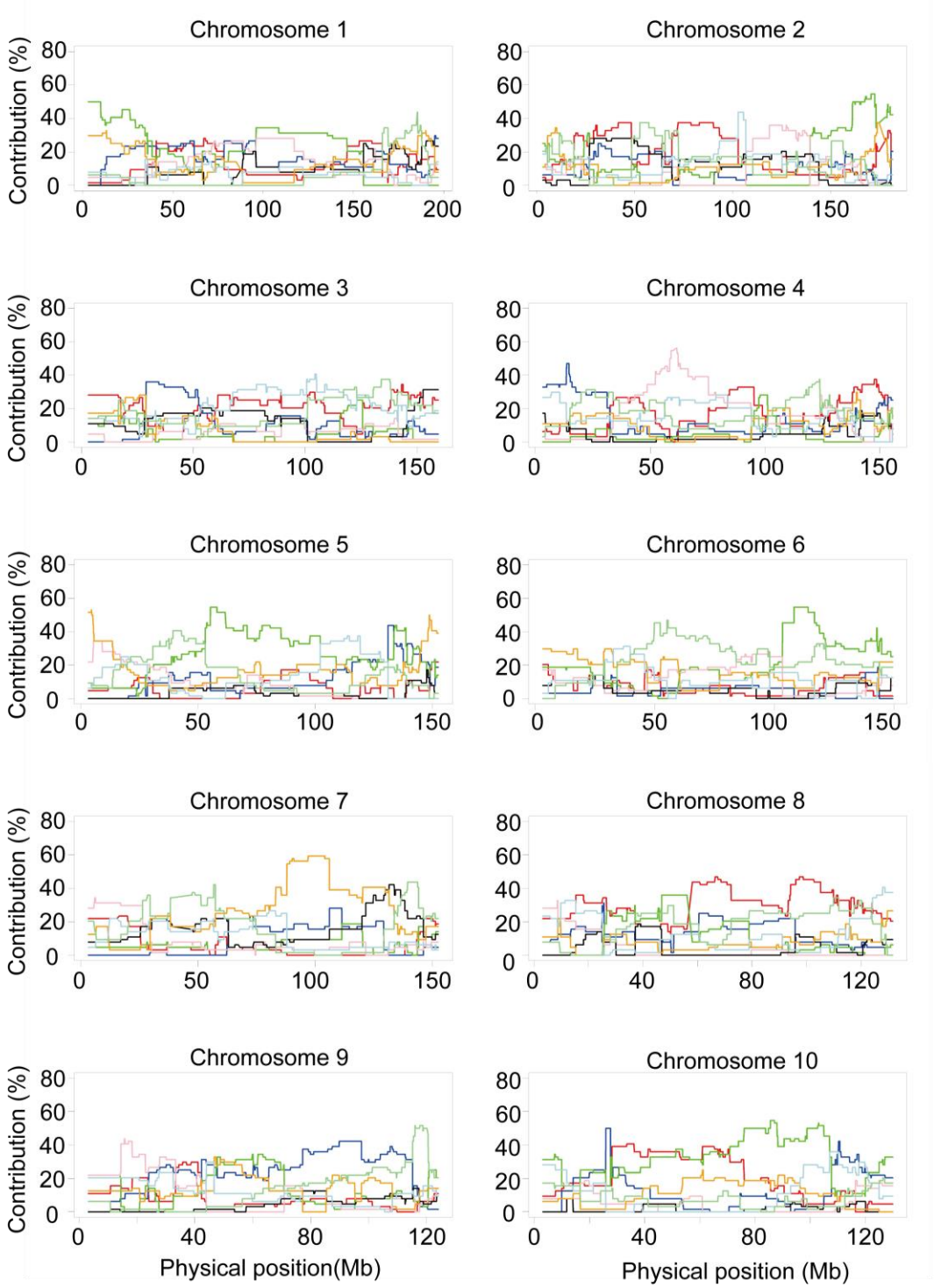

Figure 9. Contribution of each founder strain in the S2 population at G_{12}.

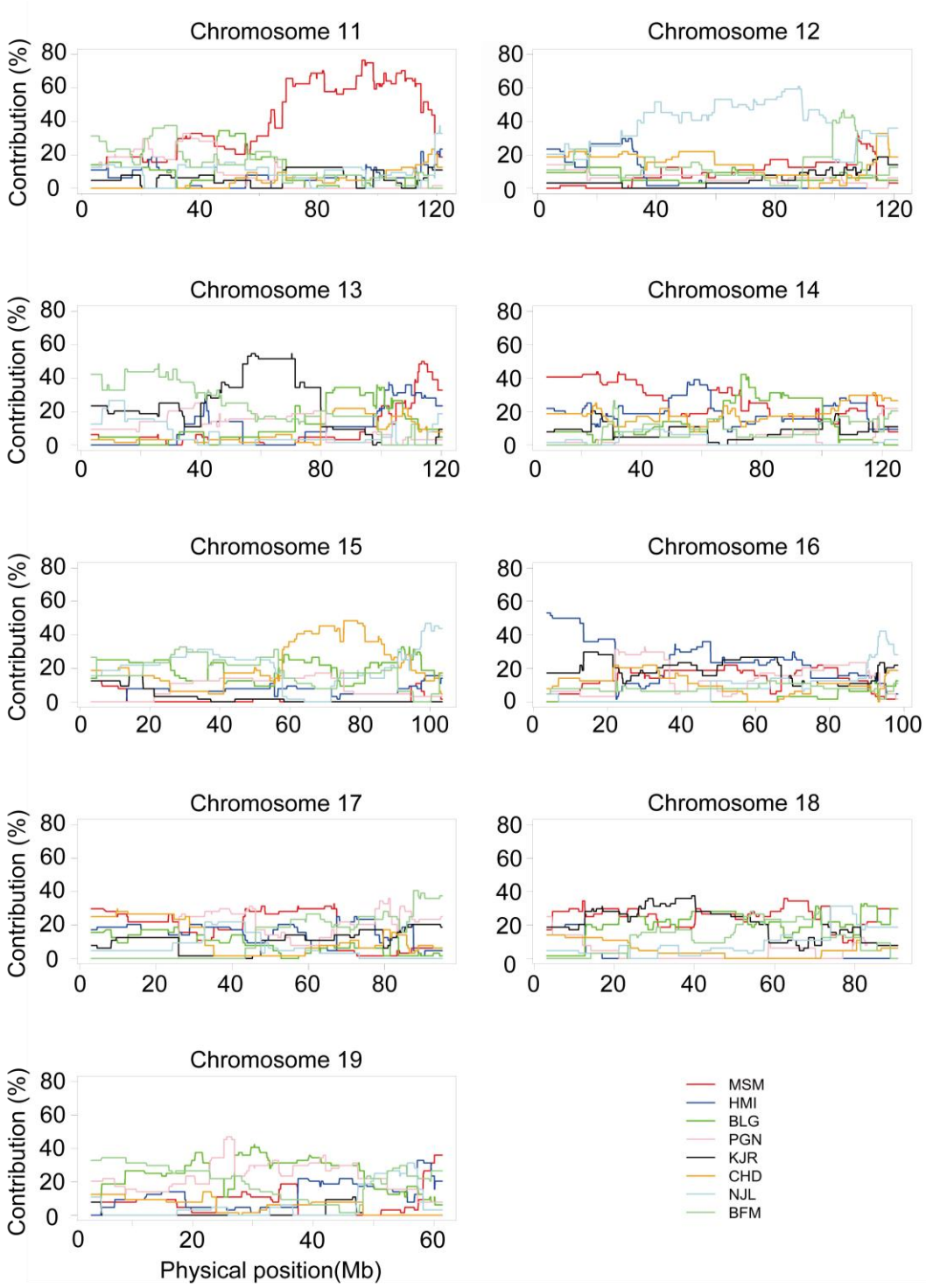

Figure 9. Continued.

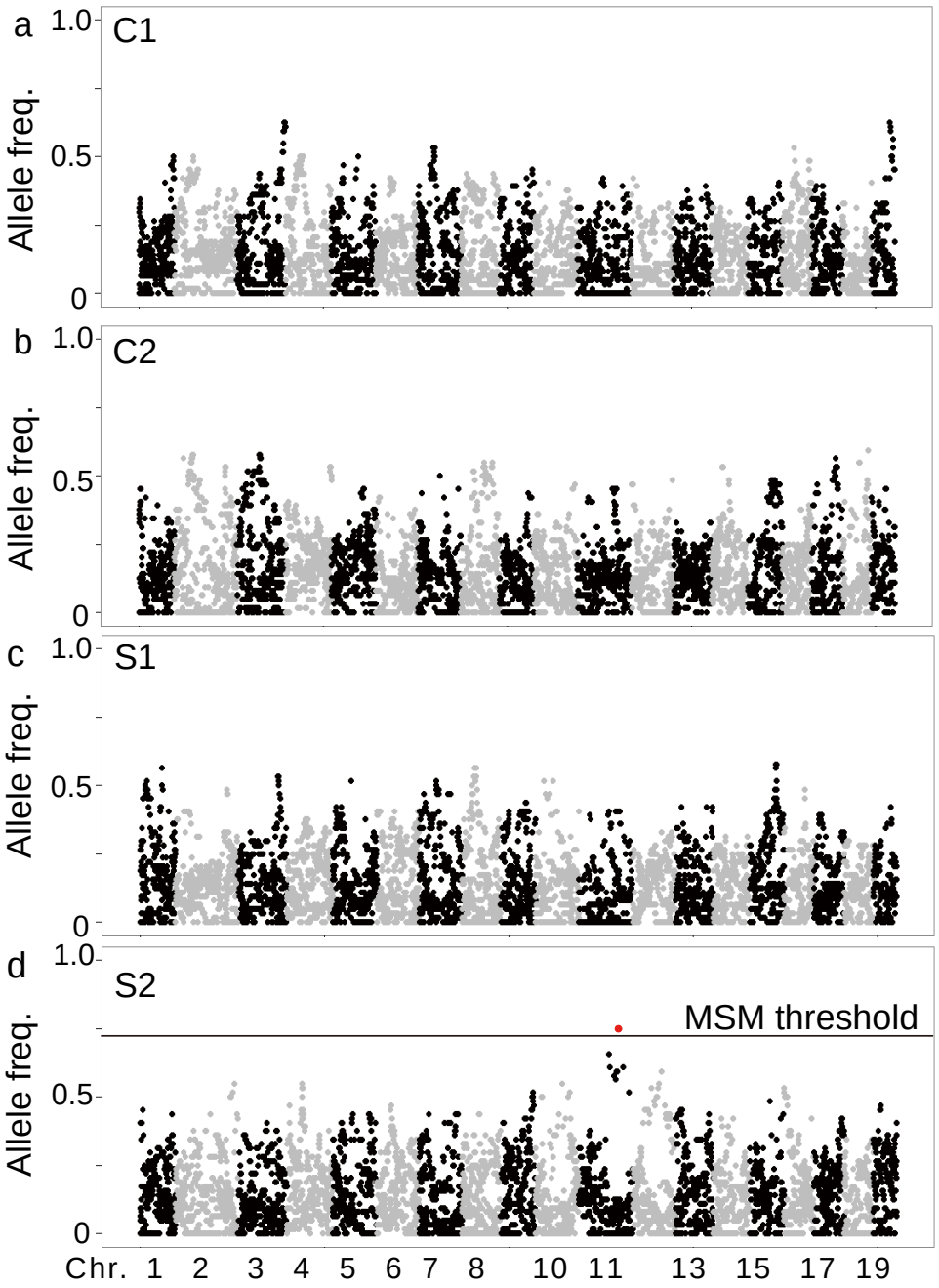

Figure 10. The allele frequency for 20,530 one-strain-specific SNPs at G_{12}. In the S1 groups, no SNP reached the threshold determined by computer simulation described in Table 5. In the S2 group, one SNP (UNC20197962) exceeded the threshold value for MSM (red dot).

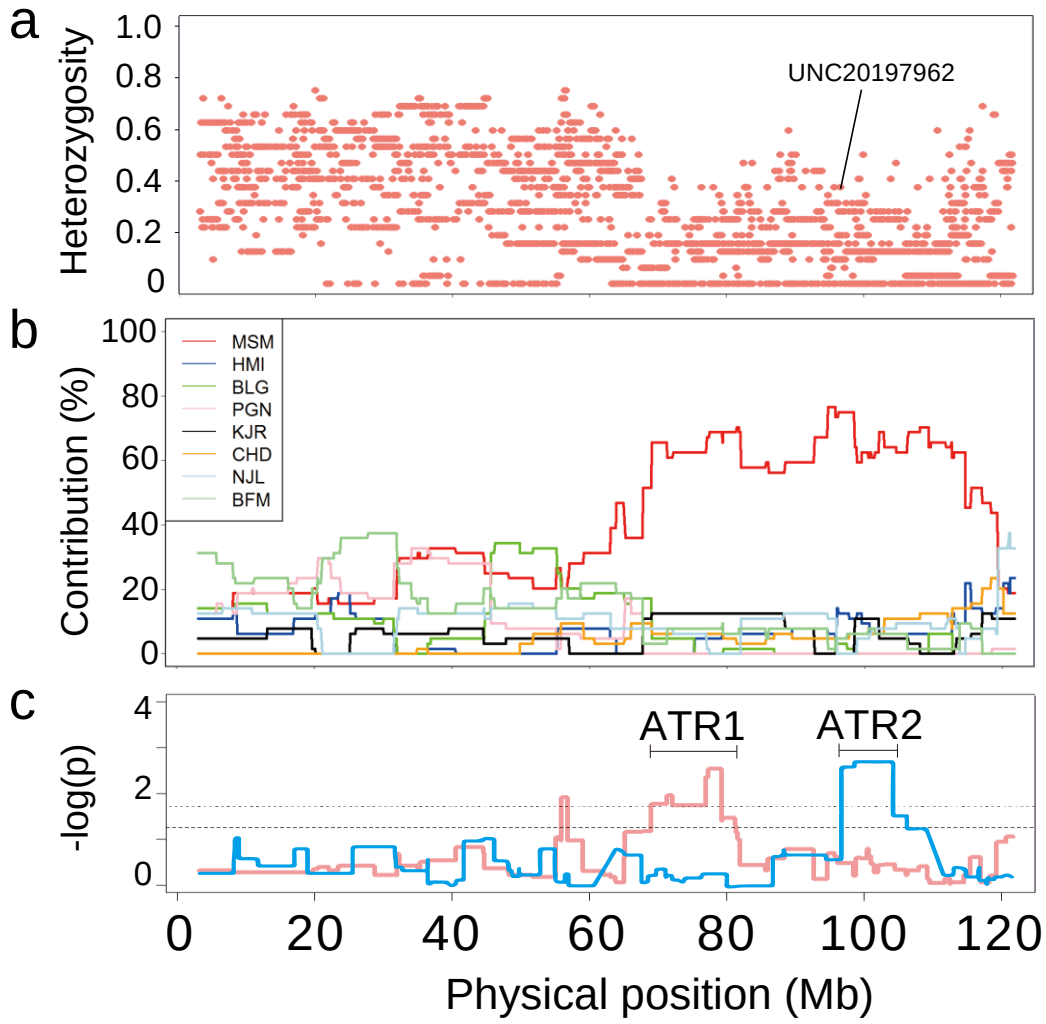

Figure 11. The haplotype derived from the MSM strain was selected in the S2 group at G_{12}. a: Heterozygosity. b: Contribution of inferred haplotype from the eight founder strains. c: Result of association between contacting and inferred MSM haplotype in S2 (pink) and C1 (blue). The selected MSM allele between 69.0 and 81.2 Mb, and 96.8 and 104.4 Mb were named ATR1 and ATR2. The dot lines indicated the thresholds for S2 (lower line) and C1 (upper line).

Table 1. The origins of WHS founder strains.

Strain	Abbr.	Subspecies group	Place of collection	Year of capture
BFM/2Ms	BFM/2	*domesticus*	Montpellier, France	1976
PGN2/Ms	PGN2	*domesticus*	Ontario, Canada	1979
HMI/Ms	HMI	*castaneus*	Heimei, Taiwan	1986
BLG2/Ms	BLG2	*musculus*	General Toshevo, Bulgaria	1980
CHD/Ms	CHD	*musculus*	Chendu, China	1981
KJR/Ms	KJR	*musculus*	Kojuri, Korea	1984
MSM/Ms	MSM	*musculus*	Mishima, Japan	1978
NJL/Ms	NJL	*musculus*	Northern Jutland, Denmark	1980

Table 2. The results of Steel–Dwass tests between two (G_3–G_4) and four populations (G_5–G_{12}).

Compared group	Gener-ation	Active tame test				Passive tame test				Stay on hand test
		Heading	Contacting	Locomotion	Jumping	Heading	Accepting	Locomotion	Jumping	Staying
C1 vs S1	G3	0.975	0.637	0.612	0.002	0.714	0.879	0.199	0.190	0.127
	G4	0.666	0.717	0.664	0.382	0.058	0.289	0.050	0.849	0.113
	G5	0.005	0.056	1.000	0.511	0.007	0.823	0.973	0.957	0.954
	G6	0.001	0.000	0.984	0.961	0.000	0.509	0.767	0.045	0.000
	G7	0.000	0.000	0.009	0.715	0.012	0.031	0.333	0.879	0.053
	G8	0.149	0.955	0.000	0.262	0.344	0.004	0.000	0.282	0.003
	G9	0.000	0.000	0.001	0.965	0.015	0.000	0.472	0.709	0.001
	G10	0.000	0.000	0.992	0.523	0.000	0.000	0.000	0.000	0.000
	G11	0.001	0.001	0.990	0.004	0.000	0.024	0.006	0.000	0.006
	G12	0.000	0.000	0.267	0.006	0.000	0.000	0.000	0.000	0.000
C1 vs S2	G5	0.775	0.507	0.878	0.094	0.283	0.706	0.388	0.507	0.523
	G6	0.767	0.006	0.994	1.000	0.517	0.654	0.984	0.916	0.113
	G7	0.010	0.000	1.000	0.836	0.172	0.622	0.146	0.963	0.689
	G8	0.005	0.030	0.000	0.875	0.321	0.403	0.019	1.000	0.117
	G9	0.001	0.000	0.425	0.717	0.000	0.000	0.035	0.057	0.000
	G10	0.000	0.000	0.791	0.929	0.000	0.004	0.000	0.001	0.000
	G11	0.000	0.000	0.963	0.463	0.000	0.000	0.000	0.000	0.000
	G12	0.000	0.000	0.016	0.355	0.000	0.000	0.000	0.000	0.000
S1 vs C2	G5	0.223	0.608	0.847	0.509	0.855	0.951	1.000	0.702	0.995
	G6	0.000	0.000	0.000	0.855	0.000	0.000	0.752	0.649	0.082
	G7	0.000	0.000	0.000	0.362	0.000	0.000	0.000	0.471	0.000
	G8	0.914	0.060	0.000	0.002	0.000	0.057	0.336	0.000	0.000
	G9	0.010	0.000	0.971	0.034	0.000	0.000	0.206	0.000	0.000
	G10	0.002	0.000	0.828	0.000	0.000	0.000	0.000	0.000	0.000
	G11	0.000	0.000	0.347	0.000	0.000	0.000	0.005	0.000	0.000
	G12	0.036	0.000	0.666	0.000	0.000	0.000	0.000	0.000	0.000
S2 vs C2	G5	0.949	1.000	0.441	0.988	0.989	0.203	0.711	0.998	0.755
	G6	0.003	0.436	0.000	0.995	0.002	0.000	0.342	0.147	0.990
	G7	0.000	0.000	0.650	0.485	0.000	0.005	0.000	0.674	0.028
	G8	0.107	0.000	0.273	0.734	0.000	0.774	0.006	0.008	0.005
	G9	0.084	0.000	0.171	0.009	0.000	0.000	0.000	0.000	0.000
	G10	0.000	0.000	0.489	0.106	0.000	0.001	0.000	0.000	0.000
	G11	0.000	0.000	0.069	0.000	0.000	0.000	0.000	0.000	0.000
	G12	0.000	0.000	0.999	0.000	0.000	0.000	0.000	0.000	0.000
C1 vs C2	G5	0.485	0.500	0.863	0.048	0.099	0.613	0.972	0.410	0.994
	G6	0.056	0.302	0.000	0.989	0.058	0.000	0.105	0.476	0.285
	G7	0.497	0.003	0.826	0.951	0.006	0.000	0.000	0.924	0.260
	G8	0.535	0.228	0.000	0.267	0.055	0.964	0.000	0.003	0.902
	G9	0.260	0.927	0.004	0.141	0.086	0.334	0.001	0.000	0.031
	G10	0.906	0.949	0.967	0.022	0.996	0.958	0.026	0.002	0.267
	G11	1.000	0.011	0.221	0.009	0.805	0.000	1.000	0.087	0.000
	G12	0.016	0.455	0.032	0.000	0.019	0.001	0.095	0.000	0.011
S1 vs S2	G5	0.061	0.674	0.895	0.688	0.676	0.314	0.650	0.797	0.814
	G6	0.030	0.002	0.997	0.950	0.002	0.999	0.919	0.007	0.102
	G7	0.006	0.144	0.006	1.000	0.877	0.000	0.885	0.998	0.616
	G8	0.360	0.075	0.040	0.050	0.999	0.363	0.000	0.288	0.818
	G9	0.914	0.228	0.059	0.920	0.107	0.023	0.001	0.408	0.986
	G10	0.720	0.005	0.883	0.204	1.000	0.000	0.978	0.993	0.898
	G11	0.338	0.000	0.820	0.216	0.017	0.445	0.003	0.200	0.037
	G12	0.001	0.000	0.567	0.394	0.051	0.771	0.015	0.618	0.427

P-value
1.000
0.500
0.050
0.005
0.000

Table 3. Spearman rank correlation coefficient between contacting and other behavioral indices.

Test	Index	S	P value	ρ
Active tame test	heading	395088	0.3646	0.078
	locomotion	285810	0.0001	0.333
	jumping	487490	0.1089	−0.138
Passive tame test	heading	336570	0.0118	0.215
	locomotion	535556	0.0033	−0.250
	accepting	432171	0.9216	−0.008
	jumping	550369	0.0008	−0.284
Stay-on hand test	staying	322693	0.0036	0.247

Table 4. Strain differences in SNPs from the GigaMUGA array.

Strain	Subspecies	# of SNPs	%
PGN2	*domesticus*	8,109	15.6
BFM/2	*domesticus*	6,038	11.6
HMI	*castaneus*	4,533	8.7
NJL	*musculus*	724	1.4
BLG2	*musculus*	585	1.1
MSM	*musculus*	275	0.5
CHD	*musculus*	158	0.3
KJR	*musculus*	108	0.2
Total		20,530	39.4

Percentage indicates the ratio of the number of strain specific SNPs for each strain to the total number of SNP (52,135SNP).

Table 5. Thresholds for each strain-specific SNP after Bonferroni correction for eight strains.

Strain		PGN	BFM2	HMI	BLG2	CHD	NJL	MSM	KJR
# of SNPs		8,109	6,038	4,533	724	585	275	158	108
Group	C1	0.797	0.797	0.797	0.750	0.734	0.766	0.750	0.719
	C2	0.813	0.797	0.797	0.781	0.750	0.781	0.750	0.734
	S1	0.781	0.781	0.781	0.750	0.719	0.750	0.734	0.703
	S2	0.781	0.797	0.781	0.766	0.750	0.766	0.734	0.734

Table 6. Profiles of 27 genes within ATRs using keywords related to tameness. The gene information was obtained from Gene Expression Database in Mouse Genomics Informatics.

ATRs	Start	End	Symbol	Keywords	Number of keywords	Expression Enbryonic head	Postnatal brain
ATR1	69632990	69653297	*Fxr2*	anxiety, fear	2	+	+
	69823122	69837784	*Nlgn2*	anxiety, fear, novelty, social, thigmotaxis	5	ND	ND
	70017085	70045532	*Dlg4*	anxiety, approach, exploration, fear	4	+	+
	70614883	70619216	*Chrne*	exploratory	1	ND	ND
	71749920	71789647	*Wscd1*	exploration, novelty	2	+	ND
	73019008	73042073	*Camkk1*	fear	1	+	+
	73160421	73172685	*P2rx5*	exploration, exploratory, fear, novelty	4	+	ND
	73183596	73199042	*Ctns*	anxiety, fear, thigmotaxis	3	ND	ND
	73234292	73261242	*Trpv1*	anxiety, exploration, fear	3	+	+
	73304992	73329596	*Aspa*	anxiety, fear	2	+	+
	74906509	74925948	*Srr*	anxiety, fear, social, thigmotaxis	4	+	+
	76998603	77032340	*Slc6a4*	anxiety, approach, exploration, fear, novelty, social	6	+	+
	77493562	77507786	*Git1*	anxiety, fear	2	ND	ND
	78166106	78176675	*Nek8*	social	1	ND	ND
	79013440	79146407	*Ksr1*	fear	1	+	+
	79339693	79581612	*Nf1*	exploratory, fear	2	+	+
	80477023	80481184	*Cdk5r1*	anxiety, fear, exploration	3	+	+
ATR2	97205842	97280638	*Npepps*	anxiety, exploration, fear, social, thigmotaxis	5	+	ND
	97509340	97576186	*Srcin1*	exploration	1	ND	ND
	98740638	98769006	*Thra*	anxiety, fear, exploratory, social	4	+	+
	100761069	100762931	*Hcrt*	social	1	+	+
	101070012	101077672	*Naglu*	anxiety, fear	2	−	ND
	101078411	101080527	*Hsd17b1*	exploration, novelty	2	ND	ND
	102145513	102149477	*Nags*	social	1	−	ND
	102430315	102437048	*Grn*	aggression, anxiety, fear, exploration, novelty, social, thigmotaxis	7	+	+
	104132855	104175523	*Crhr1*	anxiety, exploration, exploratory, fear, social	5	+	ND
	104231390	104332090	*Mapt*	anxiety, approach, fear	3	+	+

ND, no data was registered. "−", expression was not detected. "+", expression was registered in the Gene Expression Database. The gene information was obtained from Gene Expression Database in Mouse Genomics Informatic

Table 7. Overlapped regions of ATRs in mice with tameness QTL in rats and selected region during the dog domestication.

Location ID	Mouse				Compared animals					Overlap Probability
	Region	Chr.	Start (bp)	End (bp)	Species	Chr.	Start (bp)	End (bp)	Reference	
1	ATR2	11	100,887,204	104,407,876	Rats	10	90,200,000	93,721,929	Heyne et al. 2014	0.214
2	ATR1	11	76,837,858	76,892,608	Dogs and wolves	9	47,480,000	47,580,000	Wang et al. 2013	0.051
3	ATR1	11	80,210,764	80,293,643	Dogs and wolves	9	43,820,000	43,920,000	Wang et al. 2013	-
4	ATR2	11	100,436,282	101,500,066	10 dog breeds	9	23,182,777	24,182,777	Akay et al. 2010	0.096