








Reconstructing the Evolutionary History of Chromosomal Races on Islands: A Genome-Wide Analysis of Natural House Mouse Populations

Paolo Franchini ^{*,1} Andreas F. Kautt ^{1,2} Alexander Nater ¹ Gloria Antonini ³ Riccardo Castiglia ³
Axel Meyer ¹ and Emanuela Solano ³

¹Department of Biology, University of Konstanz, Konstanz, Germany

²Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA

³Department of Biology and Biotechnology "Charles Darwin," "La Sapienza" University of Rome, Rome, Italy

Associate editor: Rebekah Rogers

*Corresponding author: E-mail: paolo.franchini@uni-konstanz.de.

Abstract

Chromosomal evolution is widely considered to be an important driver of speciation, as karyotypic reorganization can bring about the establishment of reproductive barriers between incipient species. One textbook example for genetic mechanisms of speciation are large-scale chromosomal rearrangements such as Robertsonian (Rb) fusions, a common class of structural variants that can drastically change the recombination landscape by suppressing crossing-over and influence gene expression by altering regulatory networks. Here, we explore the population structure and demographic patterns of a well-known house mouse Rb system in the Aeolian archipelago in Southern Italy using genome-wide data. By analyzing chromosomal regions characterized by different levels of recombination, we trace the evolutionary history of a set of Rb chromosomes occurring in different geographical locations and test whether chromosomal fusions have a single shared origin or occurred multiple times. Using a combination of phylogenetic and population genetic approaches, we find support for multiple, independent origins of three focal Rb chromosomes. The elucidation of the demographic patterns of the mouse populations within the Aeolian archipelago shows that an interplay between fixation of newly formed Rb chromosomes and hybridization events has contributed to shaping their current karyotypic distribution. Overall, our results illustrate that chromosome structure is much more dynamic than anticipated and emphasize the importance of large-scale chromosomal translocations in speciation.

Key words: chromosomal speciation, demographic history, differential gene flow, ddRAD, Robertsonian races, zonal raiation.

Introduction

The role and impact of chromosomal rearrangements (inversions, deletions, duplications, and translocations) in speciation continues to be debated in evolutionary biology (Coyne and Orr 2004; Faria and Navarro 2010; Kirkpatrick 2017). Species often differ substantially in their karyotypes, and extensive chromosomal rearrangements are frequently observed even between closely related taxa (White 1973; King 1993). This has led to several hypotheses on how such structural changes may be driving speciation. For instance, these changes could lead to hypofertility of hybrids that is related to the underdominant nature of these mutations. They could also favor reproductive isolation due to the initial localized reduction in both recombination rates and gene flow between diverging populations (Bush et al. 1977; Leache et al. 2016). However, many chromosomal rearrangements seem to have little effect on fertility and structural variants between species may accumulate only after the establishment of

reproductive isolation rather than causing it (reviewed by Rieseberg [2001] and Faria and Navarro [2010]).

Robertsonian (Rb) fusions and fissions are a prominent class of large-scale chromosomal translocations and are likely the most common rearrangements in animals (King 1993). They occur in ~0.1% of human meiotic divisions (Evans et al. 1982; Song et al. 2016) and are also frequent in sheep and cattle (Pagacova et al. 2011). Rb translocations involve the physical joining of two acrocentric chromosomes at their centromeres and the subsequent formation of a new metacentric chromosome. These fusions therefore lead to a karyotypic reorganization that produces a change in diploid number, but not in genome content or number of chromosome arms (Garagna et al. 2001). New molecular tools and powerful bioinformatic algorithms now make it possible to gain detailed insights into the links between structural variation and speciation. Rb rearrangements are key candidates. For example, a recent phylogenomic study on a species-rich

genus of lizards, *Sceloporus*, revealed higher speciation rates in clades with extensive Rb fusions (Leache et al. 2016).

Rb fusions are known to occur in high proportions in natural populations of the Western European house mouse (*Mus musculus domesticus*), making it an ideal model for addressing questions on chromosomal speciation. To date, more than 100 populations with metacentric chromosomes have been discovered (Pialek et al. 2005; Hauffe et al. 2011; Castiglia et al. 2015), and in many of them Rb fusions are fixed and form the so-called “chromosomal races” with diploid numbers ranging from $2n = 40$ (all acrocentric; standard karyotype of the genus *Mus*) to $2n = 22$ (up to nine metacentrics; highest number of fused chromosomes).

How do new chromosomal populations and chromosomal races arise? How easily do newly formed metacentric chromosomes become fixed in a population despite their potential negative effects on fitness? Rb fusions can alter the recombination landscape either 1) because the frequency of meiotic chiasmata is reduced in genomic regions linked to the translocation, both in homo- and hetero-karyotypes (Bidau et al. 2001; Castiglia and Capanna 2002; Dumas and Britton-Davidian 2002) or 2) because of aneuploidy of recombinant gametes that suffer strong fitness costs (underdominance). Indirect evidence for “recombination suppression” comes from genetic surveys targeting hybridization areas between Rb races characterized by different karyotypes. High genetic divergence was mainly localized in regions close to the centromeres of Rb chromosomes due to their lower permeability to gene flow (Panithanarak et al. 2004; Franchini et al. 2010; Forster et al. 2013, 2016). The localized suppression of recombination might promote adaptive evolution by building up linkage disequilibrium between advantageous allelic combinations and therefore facilitate the fixation of chromosome changes despite reduced fertility of heterokaryotypes. Such rapid fixation of Rb chromosomes can be further promoted by meiotic drive or simply by beneficial effects of the altered chromosomal organization on gene expression (Franchini et al. 2016; Lindholm et al. 2016). Together with these selective processes, genetic drift may also be important in the fixation of new chromosomal rearrangements, especially when populations are small, as is often the case in the house mouse; a species mainly commensal with humans that commonly live in isolated rural areas. New Rb races can also arise without the formation and subsequent fixation of new Rb chromosomes, simply by exchanging existing Rbs through hybridization. The newly developed Rb combinations formed via this process, known as “zonal raiation” (Searle 1993; White et al. 2010), can face different fates and will either go extinct or become fixed, depending on the prevailing evolutionary forces acting on them.

Given the potential role of these large-scale karyotypic rearrangements in promoting speciation, determining the rate at which Rb chromosomes arise can indirectly inform us about the rate at which house mice natural populations might establish reproductive isolation (Riginos and Nachman 1999). Estimating the evolutionary history of Rb races is an indirect way of measuring this rate, where evidence for the “multiple-independent origin” hypothesis (i.e., the

independent origin of the same Rb fusions in geographically separated populations) would support a high mutation rate and, therefore, indirectly a potentially repeated and active role of Rb translocations in speciation. Conversely, support for a “single-origin” hypothesis (i.e., most Rbs derive from a single lineage followed by long-distance dispersal) would speak for a lower mutation rate and therefore less of a chance that they may be involved in speciation (at least repeatedly). Recent studies suggest that a combination of these two mechanisms may have influenced the current distribution of Rb populations of house mice (Gimenez et al. 2016); however, we have yet to understand the relative contributions of single versus multiple origins in shaping current populations.

To this end, we explored the evolutionary history of the Aeolian islands’ Rb system using genome-wide double-digest restriction site-associated DNA (ddRAD) markers. The Aeolian archipelago, located off the shore of Southern Italy (fig. 1), is composed of seven small islands of volcanic origin and is inhabited by populations with both Rb and standard karyotypes (fig. 1). Previous research on this system was instrumental in characterizing the karyotypes occurring on the islands (Solano et al. 2007, 2009). However, attempts to elucidate the evolutionary history of the mouse populations using cytogenetics and mitochondrial DNA have so far been inconclusive. Genome-wide markers have the potential to provide the resolution required for inferring evolutionary histories related to fusion events by enabling the accurate quantification of population genetic parameters along each chromosome. In particular, the centromeric-linked regions, characterized by low recombination rate, have been shown to be effective in depicting the evolutionary origin of Rb chromosomes (Riginos and Nachman 1999; Forster et al. 2013; Gimenez et al. 2016).

In this study, we investigate the population structure of the house mouse Rb system in the Aeolian islands and the demographic history of the focal mouse populations. We test whether a set of three metacentric chromosomes [Rb(1.2), Rb(3.9), and Rb(6.16)], found in the Aeolian islands, were introduced to the islands from a population in Central Italy as previously hypothesized, or if they originated on the islands in situ (fig. 1). This permitted to reconstruct the colonization routes of the mice within the archipelago and to disentangle the relative contributions of the fixation of newly formed Rb chromosomes versus introgression/hybridization with the current distribution of rearranged karyotypes.

Results

Population Structure

All analyses investigating genetic differentiation in our data set suggested well-defined population structure with no evidence for recent gene flow (fig. 2). SplitsTree identified a network in which all individuals of the eight populations form clearly distinguishable monophyletic groups with no evident reticulation (fig. 2a). This pattern was confirmed by an Admixture analysis. When this Bayesian algorithm was set to eight genetic clusters, corresponding to the number of

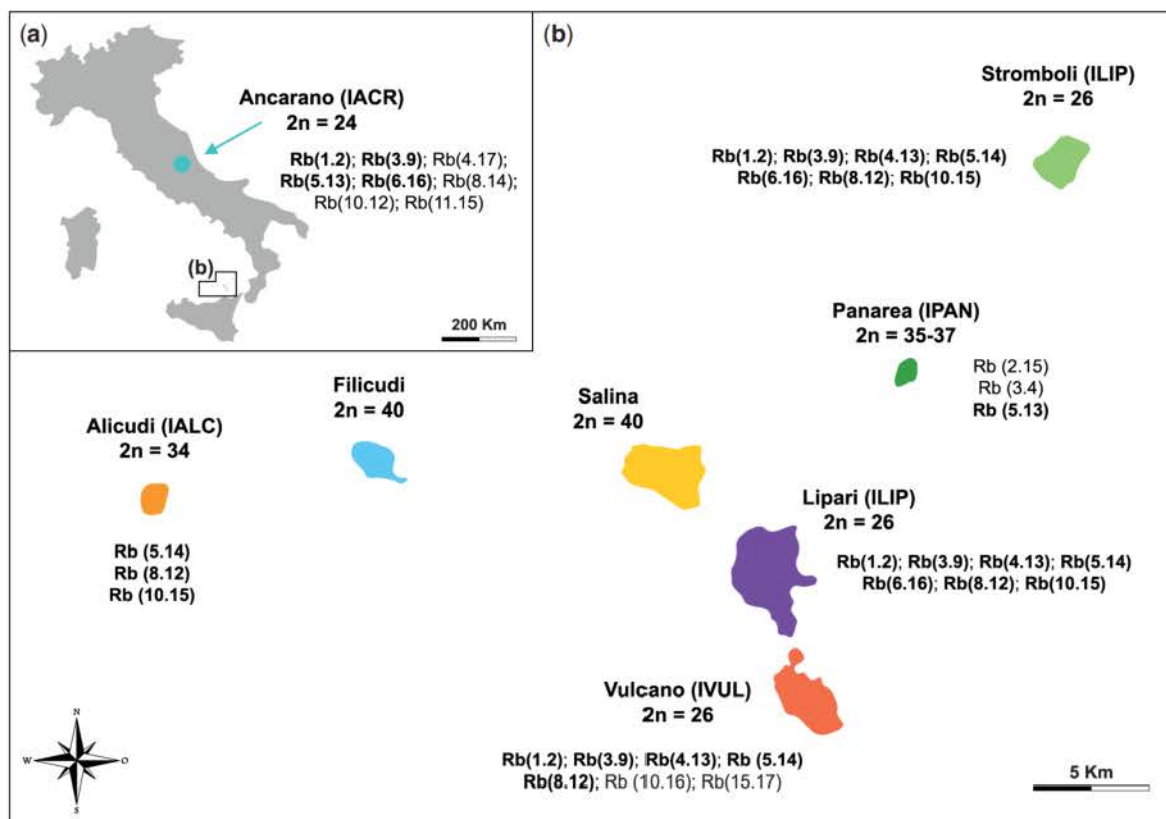


FIG. 1. Karyotypes of the eight house mouse populations analyzed in this study. (a) Map of Italy showing the mainland Ancarano population and the location of the Aeolian islands, displayed at a larger scale in (b). Robertsonian (Rb) chromosomes are highlighted in bold when they are shared by at least two populations. Robertsonian chromosomal races' names are in parentheses. Number of individuals per locality are as follows: Ancarano, 12; Alicudi, 14; Filicudi, 14; Lipari, 12; Panarea, 4; Salina, 7; Stromboli, 10; and Vulcano, 16.

islands plus the mainland ($K = 8$), all individuals were assigned to the correct population of origin with very weak signatures of admixture in only three Lipari samples (fig. 2c). Other K values with high statistical support were 5, 6, 7, 9, and 10 (supplementary fig. S1, Supplementary Material online); as a general trend, $K_s < 8$ reflect a close evolutionary history of the islands collapsed in the same genetic clusters, whereas $K_s > 8$ indicate substructuring within islands that might reflect isolation by distance of different sampling localities. Eight distinct genetic clusters were also apparent in a principal component analysis (PCA) (fig. 2b). The PCA suggested that the most divergent populations were Vulcano and Ancarano, the former being clearly differentiated in the first axis of variation (PC1) and the latter in the second axis (PC2). Lipari and Stromboli, the populations sharing the same karyotype (ILIP), though forming two distinguishable groups, were very close in PCA space. Finally, F_{ST} -statistics were congruent with the pattern revealed by the clustering approaches. Generally, we found moderate to high pairwise levels of genetic differentiation across the whole data set (average $F_{ST} = 0.419$, ranging from 0.227 to 0.626), with the highest F_{ST} -values shown by Ancarano versus other populations (the average F_{ST} -values of the pairwise comparisons between Ancarano vs. all the other populations was 0.500;

all pairwise F_{ST} -values are reported in supplementary table S2, Supplementary Material online).

The monophyly of each of the eight Italian populations was unambiguously recovered by a SVDquartets phylogenetic analysis (fig. 3), as each population node received full bootstrap support (100%), except Panarea (97%). Further, high bootstrap values supported the clade formed by Lipari, Stromboli, and Vulcano (the populations with more metacentric chromosomes in the archipelago) and the clade in which these three populations are the sister group of Alicudi (95% and 85%, respectively). The monophyly of the Aeolian populations, with the exception of Panarea, was also recovered with moderate bootstrap support (66%). The inference of Panarea being a sister group to all other populations, including Ancarano from mainland Italy, is congruent with the notion that extant Panarea karyotypes likely resulted from introgression of mice from Sicily (IPAL and ICAS chromosomal races) into the Aeolian populations (Solano et al. 2009, 2013; Castiglia et al. 2015). Finally, we found lower support for a sister-group relationship between Salina and the other island populations (not considering Panarea) and between Filicudi and the metacentric populations. Salina and Filicudi are the two populations with ancestral standard karyotype that might share a more recent evolutionary history.

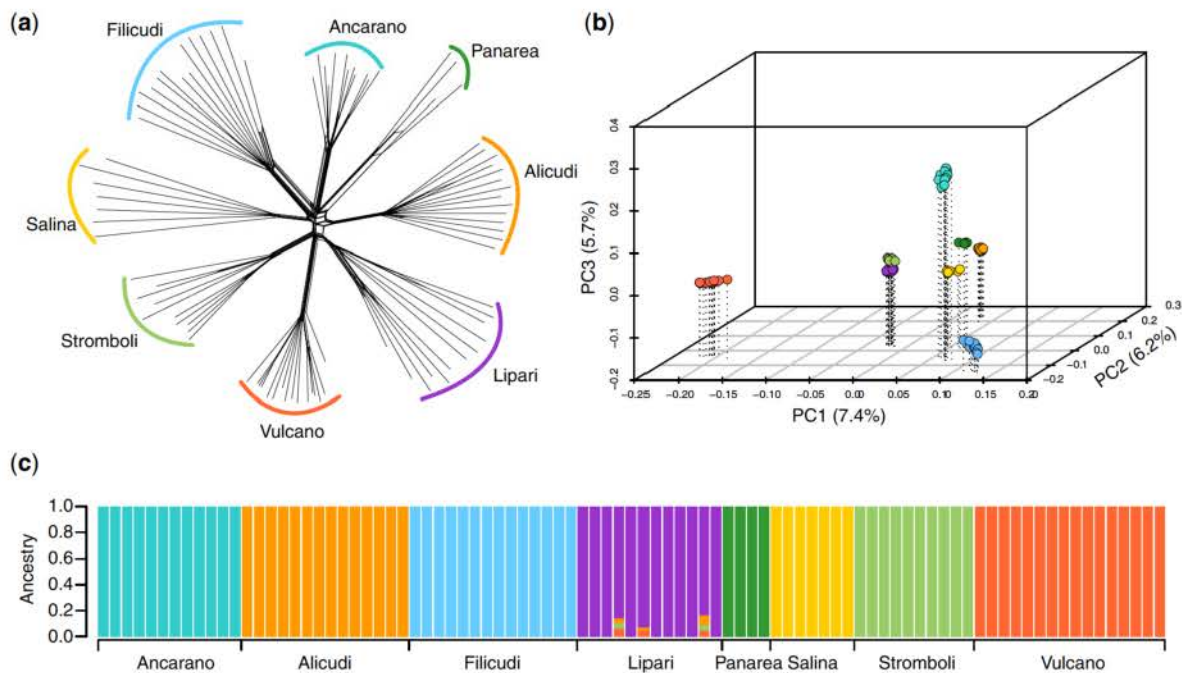


FIG. 2. The eight populations are genetically distinct with no evidence of recent gene flow. (a) An individual-based phylogenetic split network identified clearly distinguishable groups that correspond to the eight populations. (b) The three main axes of genetic variation of a PCA grouped together individuals of the same populations and identified divergent (Vulcano along PC1 and Ancarano along PC2) and genetically more closely related (Lipari and Stromboli) populations. (c) Genetic clustering and individual ancestry at $K = 8$, one of the most supported number of clusters in an Admixture analysis, revealed no evidence of recent admixture between populations. The color code is identical to figure 1.

Taken as a whole, the tree topology was consistent with Rb fusions originating in standard populations that had previously colonized the archipelago.

Potential admixture between populations was further investigated using TreeMix. It has long been acknowledged that simple bifurcating trees do not account for possible genetic exchange between populations, and thus may provide an incomplete representation of their histories. TreeMix implements a model that accounts for both splits and gene flow (Pickrell and Pritchard 2012). The inferred maximum likelihood bifurcating tree was tested by modeling up to five migration events/edges (m) on it. When we tested the robustness of the tree with no edges ($m = 0$), we find that the fit of the tree to the data was very good ($f = 0.9947$). Adding two edges ($m = 2$) resulted in the largest increase of likelihood and increased the fit slightly ($f = 0.9983$), whereas only slight likelihood changes were observed when adding additional edges ($m = 3-5$). The model with two migration edges ($m = 2$) places the first edge between Lipari (source) and Alicudi (recipient) and the second edge between Alicudi (source) and Panarea (recipient). In this latter case, the migration edge originates from the basal part of the branch leading to Alicudi, close to the Salina-Alicudi bifurcation node (fig. 4).

Linkage-Dependent Population Differentiation

Different regions on a chromosome can experience differential permeability to genetic exchange due to different

recombination rates (Charlesworth et al. 1993). Meiotic recombination is known to be reduced in regions linked to the centromeres in all eukaryotic species (Talbert and Henikoff 2010); especially when Rb translocations are involved. For this reason, loci linked to centromeres are more powerful in depicting the evolutionary relationships of populations carrying metacentric chromosomes. Here, we used two subsets of loci, linked to either centromeric or telomeric regions, to reconstruct the demographic history of the Aeolian populations.

As a first hypothesis-free exploratory analysis, we conducted PCAs with all eight populations using either one of the two subsets of loci for all autosomes. The most striking difference between the two PCAs is the position of Vulcano, which is well separated from the other populations along PC1 using telomeric loci only, whereas it clusters together with Lipari and Stromboli when using centromeric loci only (supplementary fig. S2, Supplementary Material online).

We then used linkage-dependent population differentiation approaches to explore different colonization scenarios that might explain the current karyotypes of the Aeolian mice. First, we wanted to test whether the occurrence of the three metacentric chromosomes [Rb(1.2) Rb(3.9) Rb(6.16)] found both in the archipelago (Lipari, Stromboli, and Vulcano) and on the mainland (Ancarano population) could be explained by a single origin (in Ancarano) followed by colonization of the archipelago or by multiple, independent origins of these metacentrics. If the single-origin

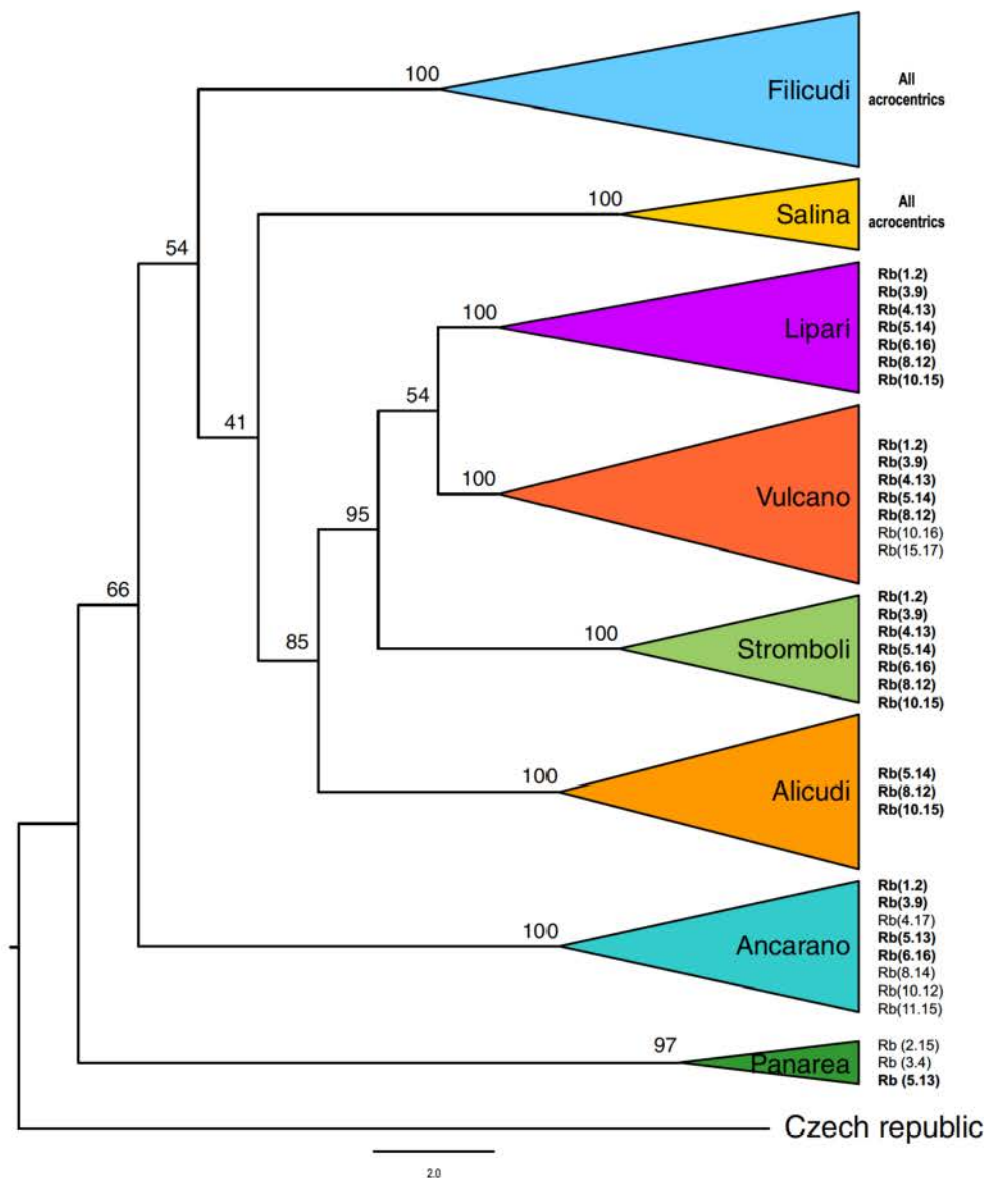


FIG. 3. Phylogenetic tree computed by the program SVDquartets on a matrix of 12,219 SNPs. The tree was rooted with a WGS individual of *Mus musculus* from the Czech Republic. Individuals from each population formed a monophyletic group and were collapsed in colored triangles, whose size reflects their number. Bootstrap support for each node is shown. Robertsonian (Rb) chromosomes characterizing the mice are reported and highlighted in bold when they are shared by at least two populations.

hypothesis was correct, we would expect to observe lower genetic differentiation between the Aeolian population and Ancarano when using centromeric than telomeric markers of such three Rb chromosomes. PCA allowed us to reject this hypothesis as the clustering of Aeolian populations and Ancarano was similar in the two different PCAs conducted on the centromeric and telomeric marker sets of the three focal metacentric chromosomes (supplementary fig. S3, Supplementary Material online). Admixture analyses confirmed this pattern, where mice of Alicudi, Lipari, and Ancarano were assigned to two genetic clusters ($K = 2$) using either [**Rb(1.2)** **Rb(3.9)** **Rb(6.16)**] or [**Rb(5.14)** **Rb(8.12)**

Rb(10.15)]. In both cases, mice from Alicudi and Lipari formed one genetic cluster and Ancarano's individuals were assigned to the second genetic cluster (supplementary fig. S4, Supplementary Material online). This finding is consistent with patterns of F_{ST} , where the centromeric loci showed even higher divergence than the telomeric loci (supplementary table S2, Supplementary Material online). Since F_{ST} is a relative measure of differentiation and might show larger values for centromeric loci due to their lower within-population diversity, we also assessed D_{XY} and D_A as absolute measures of population differentiation. As with F_{ST} , we observed generally larger values of D_A for the centromeric versus

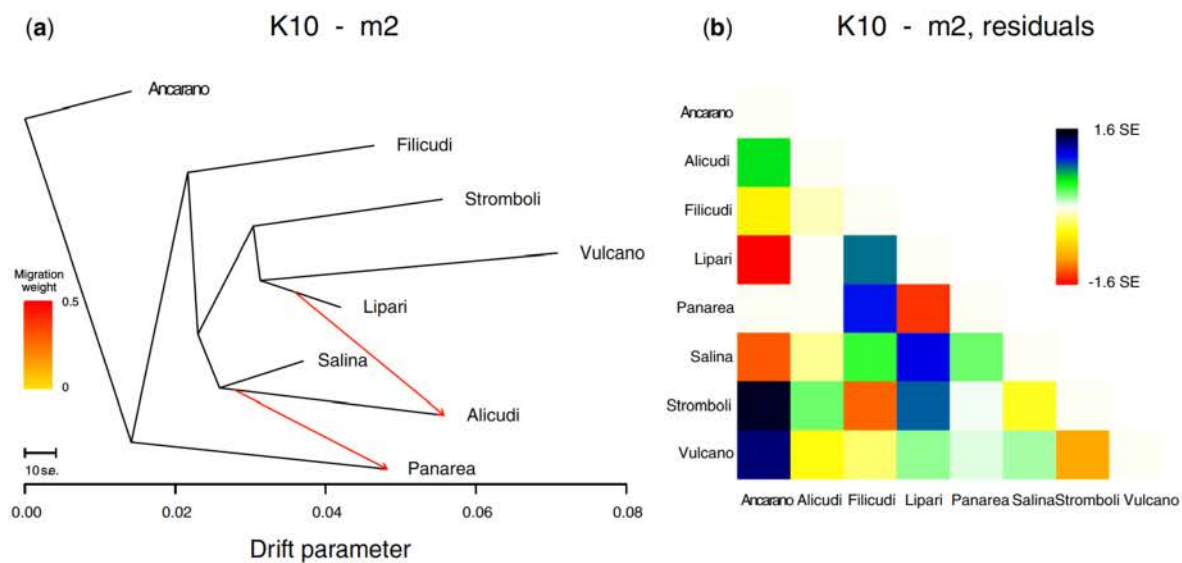


FIG. 4. TreeMix maximum likelihood tree. (a) Two migration edges ($m = 2$) have been fitted to the tree. Branch lengths reflect the amount of genetic drift. The tree was rooted with the population from Central Italy (Ancarano). The direction of gene flow is indicated by arrows and heat colors reflect intensity. (b) Heat map of the residual covariance among population pairs of the given tree on the left with two migration events.

the telomeric regions whereas more similar values of D_{XY} were found for the same two regions (fig. 5).

Second, we wanted to test whether three other metacentric chromosomes, characteristic of the archipelago [Rb(5.14) Rb(8.12) Rb(10.15)], introgressed from Alicudi into the other populations carrying today three (Lipari and Stromboli) or two (Vulcano) of them, as previously suggested. This hypothesis was rejected by the PCA results, which instead suggested an alternative scenario in which the Alicudi race could have arisen via a hybrid raiation event between mice with Lipari metacentric chromosomes and a standard population already present on the island of Alicudi. On the one hand, PCA carried out using the centromeric loci of the metacentrics shared between Alicudi and Lipari, Rb(5.14) Rb(8.12) Rb(10.15), grouped Alicudi with Lipari and Stromboli along the main axis of divergence (PC1; fig. 6b). On the other hand, Alicudi clusters with the islands inhabited by mice with a standard karyotype when PCA was conducted using the centromeric loci of the chromosomes that are Rbs in Lipari and acrocentric in Alicudi (1, 2, 3, 6, 9, and 16; fig. 6c). This finding is supported by the Admixture analyses run on subsets of populations (Alicudi, Lipari, Stromboli, Filicudi, and Salina) and chromosomes allowing K to range from 2 to 5. When admixture was restricted to two genetic groups, Alicudi clustered with Lipari and Stromboli using Rbs 5.14, 8.12, and 10.15 and with Filicudi and Salina when using chromosomes 1, 2, 3, 6, 9, and 16 (supplementary fig. S5, Supplementary Material online).

Third, using the same approach, we tested whether Vulcano island was colonized by the Lipari population and whether the current Vulcano (IVUL $2n = 26$) karyotype had arisen through a type-C whole-arm reciprocal translocation (WART) (Capanna and Redi 1995). The latter is an additional mechanism of Rb formation involving chromosomal arm

exchange between two metacentric and one acrocentric chromosome (Pialek et al. 2005). To this end, we performed independent PCAs focusing on centromeric loci that mapped to all autosome arms of the island populations (fig. 6d) and to telomeric loci of the same chromosomes and populations (fig. 6e; Panarea was excluded from this analysis because of the low sample size and uncertain origin). Additionally, to further focus on the allelic variation explained by the three focal races, we analyzed the metacentric chromosomes shared by Lipari, Stromboli, and Vulcano [Rb(1.2), Rb(3.9), Rb(4.13), Rb(5.14), and Rb(8.12)], and on the two chromosomes that are in Rb state only in Vulcano [Rb(10.16), Rb(15.17)]. The two PCAs showed very similar results, where individuals of each island formed distinguishable groups with comparable intercluster distances in the first three axes of variation (supplementary fig. S6, Supplementary Material online).

Discussion

Here, we investigated the evolutionary history of a well-known Western European house mouse Rb system from the Aeolian archipelago off the coast of Southern Italy. We used a genome-wide reduced-representation sequencing approach (ddRAD) to examine linkage-dependent, localized, genetic divergence. These genome-wide analyses led us to refute previous hypotheses suggesting that a set of three metacentric chromosomes [Rb(1.2), Rb(3.9), and Rb(6.16)], present both in the Aeolian islands and in Central Italy (Ancarano, IACR Robertsonian race), were introduced to the archipelago from the mainland population (Solano et al. 2009). Our data instead support the “multiple-independent origin hypothesis”, that is, that the same Rb fusions arose independently in geographically separated populations. Thereby, our results suggest that the origin and fixation of

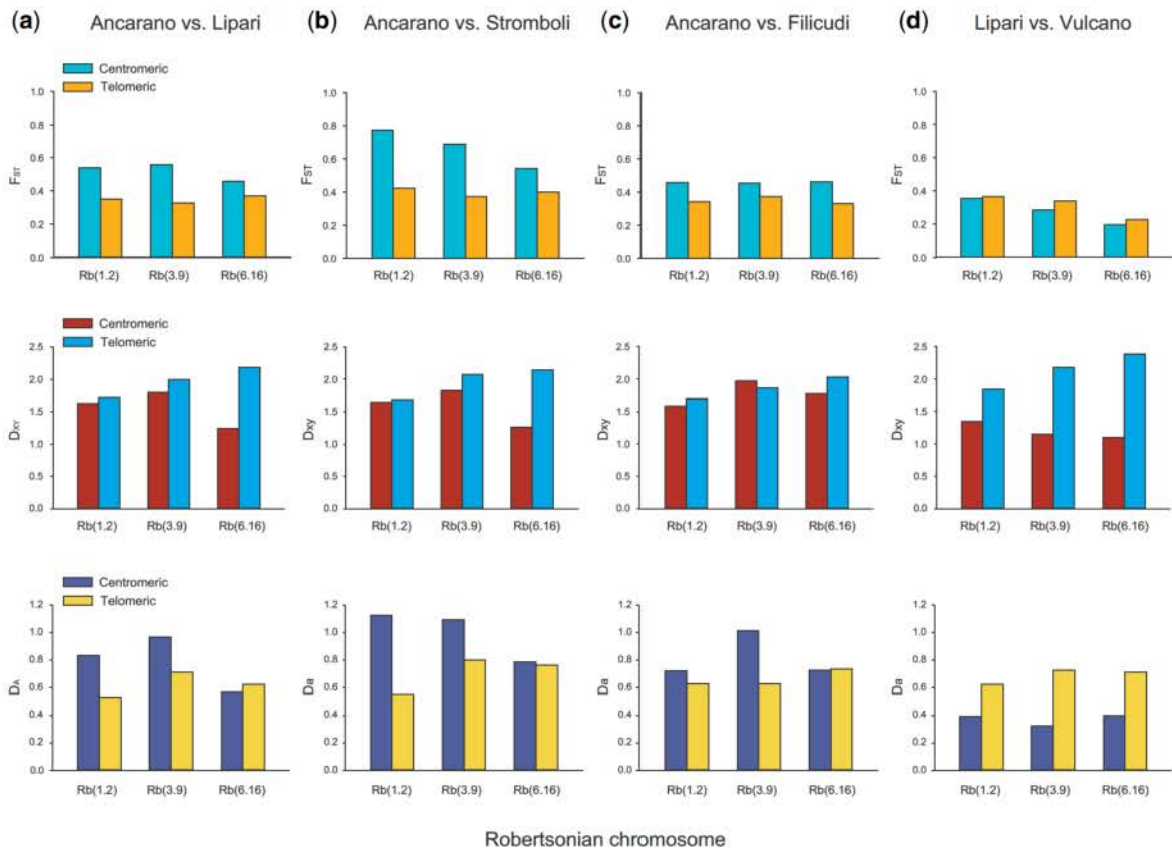


Fig. 5. “Multiple-independent origin” versus “single-origin” hypothesis. (a) F_{ST} , D_{XY} , and D_A estimated in the comparisons between Ancarano (Central Italy) and Lipari, (b) Ancarano and Stromboli, (c) Ancarano and Filicudi, and (d) Lipari and Vulcano, for both centromeric and telomeric regions of the three Robertsonian chromosomes potentially introgressed to the Islands’ populations from Central Italy. Although populations in the first two comparisons share the three focal Rb chromosomes and were used to assess the two alternative hypotheses, the third comparison served as a control (Filicudi’s mice have a $2n = 40$ all-acrocentric standard karyotype with no fusions). The fourth comparison was used as a proof-of-concept of the method demonstrating the expected patterns of a “single-origin” scenario of the three Rbs (Vulcano was colonized by mice from Lipari, clearly demonstrated by the other analyses). F_{ST} -values suggested a “multiple-independent origin” of the focal Rb chromosomes as the most likely scenario. The higher centromeric values are likely a result of background selection in genomic regions characterized by low recombination, which was also supported by lower centromeric D_{XY} -values, indicating reduced effective population size in the common ancestral populations. Net nucleotide divergence (D_A), an estimator of population split time, was in line with F_{ST} , suggesting a further reduction of recombination rates due to recent postdivergence chromosome fusions.

this potentially underdominant type of large-scale translocation might be more common than previously thought. Additionally, we inferred the population structure of the focal mice within the Aeolian archipelago and reconstructed their evolutionary history. This work highlights how fixation of newly formed Rb chromosomes, hybridization events, and absence of recent gene flow among populations could have shaped the current karyotypic distribution of the islands’ mice.

Meiotic recombination is reduced at regions linked to the centromere in all eukaryotic species (Talbert and Henikoff 2010), an effect that is further pronounced in chromosomal arms with Rb translocations (Dumas and Britton-Davidian 2002; Franchini et al. 2010). Pericentromeric regions of Rb chromosomes are more resistant to gene flow than their distal counterparts as a result of localized suppression of recombination (Panithanarak et al. 2004; Franchini et al. 2010;

Forster et al. 2013). This process was instrumental in conceiving the “recombination suppression” model, that together with the “hybrid hypofertility” model, describes the theoretical framework for explaining how Rb populations can accumulate genetic divergence and establish reproductive isolation (Faria and Navarro 2010; Potter et al. 2017). Additionally, as a consequence of the localized suppression of recombination, loci linked to the centromere are particularly useful for tracing the evolutionary history of Rb chromosomes (Riginos and Nachman 1999; Forster et al. 2013, 2016). When a Rb population colonizes a new island it may exchange genetic material with an autochthonous population already present. Similarly, source population can exchange genetic material with surrounding mainland populations. Importantly, in both cases, genetic divergence between the source population and the colonizers is expected to increase at a slower rate at pericentromeric regions of Rb

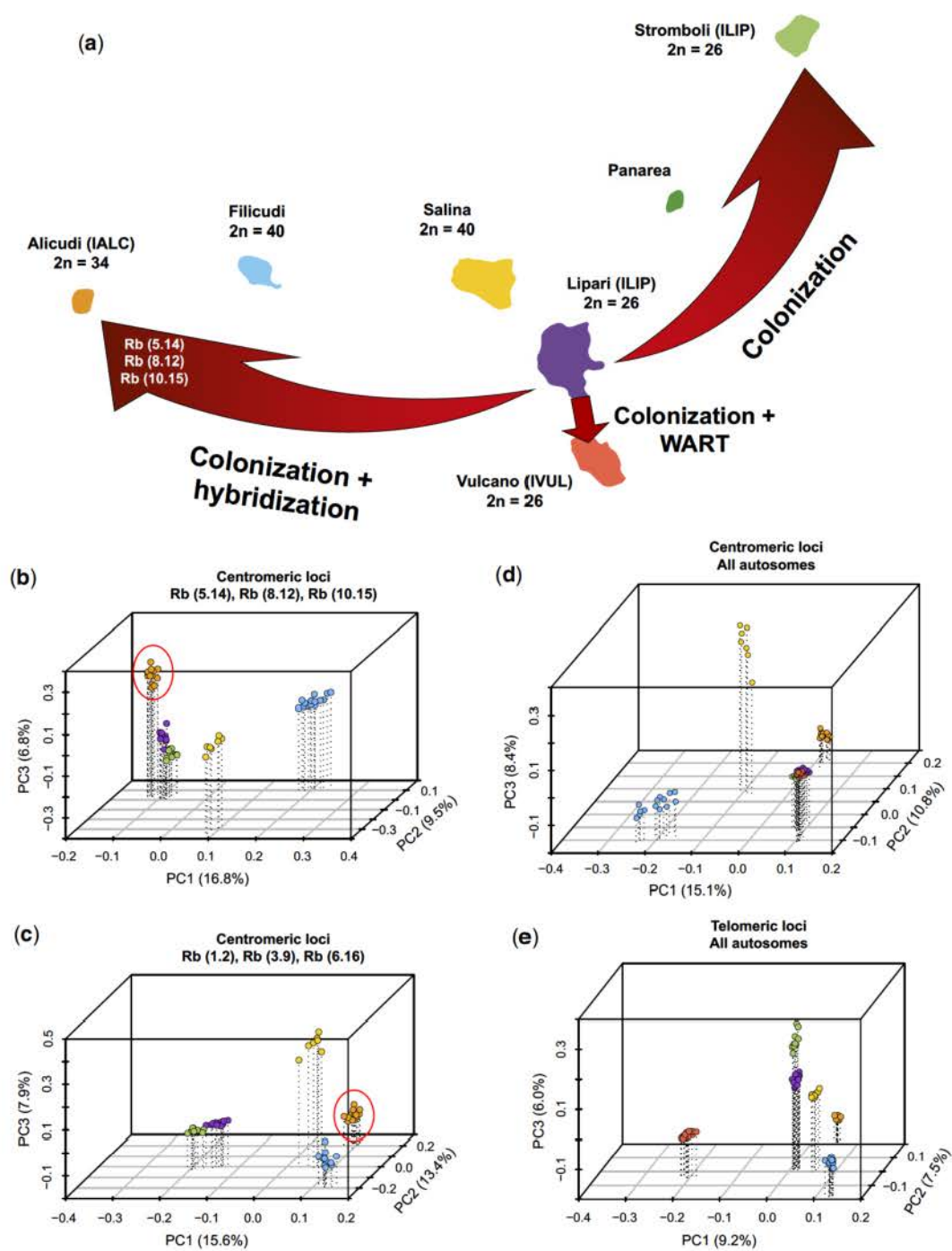


FIG. 6. Evolutionary history of the Aeolian archipelago mice. (a) Colonization scenario suggested by the genomic analyses. (b) PCA conducted on the centromeric loci of the three Rb chromosomes of the chromosomal race from Alicudi island. Mice from Alicudi (red circle) cluster closer to those from Lipari and Stromboli (sharing the three Rb fusions with Alicudi) than to Salina and Filicudi, inhabited by mice carrying the $2n = 40$ all-acrocentric standard karyotype. (c) PCA conducted on the centromeric loci of the chromosomes that are Rbs in Lipari and acrocentric in Alicudi. Mice from Alicudi cluster close to the mice with standard karyotype. (d, e) PCAs showing a different clustering of Vulcano when only centromeric loci from all autosomes (d) and only the telomeric loci (e) were used. The centromeric data set clearly shows tight clustering (i.e., shared evolutionary history) of Vulcano, Lipari, and Stromboli.

chromosomes when compared with more distal regions. The pericentromeric regions of Rb chromosomes are less likely to incorporate divergent alleles from other populations (either island's autochthonous or surrounding mainland populations) because of suppressed recombination.

In this system, if the three metacentrics [Rb(1.2), Rb(3.9), and Rb(6.16)] found in the Aeolian archipelago's mice were derived from metacentrics found on the mainland (as represented by the Ancarano population in our data set), we would consequently expect reduced divergence at the pericentromeric loci of these chromosomes in comparison to more distal chromosomal regions. However, the results we obtained using several approaches led us to reject this hypothesis. Specifically, 1) *F*-statistics detected higher relative divergence at loci linked to the translocations, an effect probably caused by reduced levels of neutral variation due to purifying selection acting on deleterious mutations in low recombination regions (background selection: Charlesworth et al. 1993; Irwin et al. 2018). The same trend might be explained by the "centromere drive hypothesis," which proposes that a centromere can act as a selfish element by increasing its own transmission through female meiosis (Henikoff et al. 2001; Rosin and Mellone 2017). This mechanism, coupled with low recombination rates, could lead to lower intrapopulation, but higher interpopulation, diversity in centromeric regions compared with telomeric regions of the chromosomes. We also found generally higher levels of net nucleotide divergence at these loci, indicating that recent postdivergence chromosome fusion events might have further reduced recombination rates and therefore nucleotide diversity in these regions as compared with the ancestral population (fig. 5). 2) Genetic clustering approaches (i.e., PCA and Admixture analysis) identified no clear pattern distinguishing centromeric and telomeric loci of the three potentially introgressed metacentric chromosomes (fig. 6). Region-specific *F*-statistics has already been applied to address such questions in other house mouse Rb systems using microsatellite loci (Riginos and Nachman 1999; Panithanarak et al. 2004; Franchini et al. 2010; Forster et al. 2013, 2016; Gimenez et al. 2016). Here, we used a dense set of genome-wide markers to accurately trace divergence along chromosomes to address this issue. In addition, it has recently been shown that patterns of relatedness formed by population structure across the genomes can vary substantially, and exploratory methods (e.g., PCA) provide a flexible new way to discover biological drivers of genetic variation (Li and Ralph 2019). Overall our analyses suggest that Rb(1.2), Rb(3.9), and Rb(6.16) have arisen independently at least twice in different geographical locations. An ancient event hypothesis, where mice from Central Italy carrying the three focal Rb fusions colonized the islands, could challenge our conclusions. However, *M. m. domesticus* is known to have entered Eastern Europe within the last 12,000 years (Cucchi et al. 2005). The westward progression of mouse populations is even more recent and occurred within the last 5,000 years following human migrations (Cucchi et al. 2005; Gabriel et al. 2010). As Rb translocations are nearly absent outside Europe and absent in other species of the

genus *Mus*, the origin of fused chromosomes in house mice is likely very recent.

In summary, our results provide indirect evidence that Rb translocations can arise and become fixed quite often and quickly in house mice. Even though the independent formation of the same three Rb chromosomes is an event with low probability, we note that not all Rb fusions have the same probability of occurring. Despite the prediction that the molecular processes governing Rb formation could occur randomly (Redi et al. 1990), small acrocentric chromosomes are less commonly involved in Rb fusions than large ones (Gazave et al. 2003). Additionally, it is reasonable to assume that natural selection can favor specific Rb fusions because of the establishment of strong linkage disequilibrium between adaptive allelic combinations. Such fused chromosomes would then have an increased probability of fixation in natural populations, while disadvantageous fusions would be eliminated by purifying selection.

The different approaches applied here identified high divergence in all pairwise comparisons and distinct populations with no evidence of recent admixture. By testing several colonization scenarios, we find that the Rb fusions originated in situ and spread across some of the islands. The metacentric chromosomes of Panarea are an exception: individuals on Panarea likely hybridized with individuals from a more recent wave of colonization, including genetic input from Salina (see e.g., PCA plots and TreeMix migration edge). Because only four individuals could be sampled from this island, the interpretation of the estimated genetic parameters needs to be taken with caution. The three Alicudi Rb chromosomes [Rb(5.14), Rb(8.12), and Rb(10.15)] are shared with Lipari and Stromboli (the two islands inhabited by mice with the same Rb karyotype). This observation, together with the potential Rb input from Central Italy, led to the previous hypothesis that these metacentrics originated in Alicudi and then introgressed into mice of Lipari giving rise to the current Rb race (Solano et al. 2009). Our results support a different scenario where Rb fusions likely originated in Lipari and then spread to the other islands. Specifically, two main lines of evidence support this alternative scenario: 1) Several of our results suggest that mice carrying Lipari/Stromboli Rb fusions colonized Alicudi island, came into contact with populations of the standard karyotype, and that hybridization between these karyotypically different mice gave rise to the current Alicudi chromosomal race. Support for this comes from the finding that the centromeric loci of the three metacentric chromosomes shared by the two islands [Rb(5.14), Rb(8.12), and Rb(10.15)] showed high genetic similarity in the Lipari versus Alicudi comparison. Alternatively, but still in line with Rb fusions originating in Lipari/Stromboli, the chromosomes that are fused in Lipari and in acrocentric configuration in Alicudi showed high genetic affinity between mice from Alicudi and Filicudi, a race with a standard all-acrocentric karyotype. The geographic proximity of Filicudi and Alicudi and their particular location in the archipelago (they branch off from the main and largest islands, Alicudi being the most remote) may have facilitated human traffic and commercial transportation between the two small islands. 2) Whether

colonizer mice arrived on Alicudi from Lipari or Stromboli (islands inhabited by mice with the same karyotype) is difficult to assess, but the relatively strong drift observed in Stromboli (see TreeMix analysis) makes the colonization of Stromboli by a small number of founders from Lipari a more likely scenario.

The hypothesis that Vulcano's chromosomal race originated from a type-C WART event from a Lipari/Stromboli karyotypic background (Solano et al. 2007) was supported by our genomic analyses. Centromeric loci of either all autosomes or a subset of these (e.g., the metacentric fusions only present in Vulcano, which allowed us to test the hypothesis of their different origin) always formed a very tight genetic cluster, including mice from Lipari, Stromboli, and Vulcano (figs. 1 and 6). Mice from the three islands were also found to belong to a monophyletic group in the phylogenetic analysis. A very strong bottleneck, suggested by the high TreeMix drift parameter, could explain why such karyotypic reorganization promoted by a WART quickly became fixed and gave rise to a new chromosomal race. On the one hand, the occurrence of WART in house mouse natural populations has been frequently considered a mechanism responsible for chromosome variability (Pialek et al. 2005). On the other hand, WART can in some cases be considered unlikely to happen, mainly because hybrids carrying newly formed metacentric chromosomes through WART have to overcome potentially significant fertility problems due to the complex structural heterozygosity that characterizes their karyotype (Hauffe and Pialek 1997). However, past observations and recent studies on meiosis progression investigating the fertility of these hybrids provide evidence for reduced, but not complete sterility (Berrios et al. 2018; Ribagorda et al. 2019). Our study supports that such molecular mechanism may have been commonly involved in the formation of Rb races in house mouse wild populations. Intriguingly, the same analyses based on telomeric loci supported a different story, where Vulcano mice are genetically more differentiated from the other populations. Strong genetic drift may have allowed allele frequencies of loci linked to these genomic regions to deviate quickly from the islands' genetic pools and can thereby possibly explain the observed pattern.

Conclusions

We show that a comprehensive genetic analysis based on a dense set of markers permitted a rigorous analysis and interpretation of the evolutionary history of a well-known Rb system of *M. m. domesticus*. Notably, our results allowed us to determine the complex patterns of chromosomal evolution in which an interplay between fixation of newly formed Rb chromosomes and hybridization events have contributed to shaping the karyotypic distribution of the islands' populations. These analyses support a new scenario of independent origins of the same Rb fusions in geographically separated populations. Assuming that hybridization between different karyotypes is associated with reduced fertility, this suggests that Rb fusions might be a more common source of reproductive isolation than previously thought. This evidence,

combined with the observed propensity of mice carrying metacentrics to form hybrid products (i.e., the formation of chromosomal races by hybridization and fixation of new combinations of Rb chromosomes), suggests that large-scale chromosomal translocations play an active role in the speciation process.

Materials and Methods

Sampling

We examined a total of 94 wild house mice (*M. m. domesticus*) from eight localities in Italy (fig. 1). Mice were collected between 2005 and 2006 from all seven islands of the Aeolian archipelago in Sicily, and between 1998 and 2000 in an Apennine valley in Central Italy (see supplementary table S1, Supplementary Material online). The samples include Rb and standard populations, whose karyotypes were characterized at the time of sampling by bone marrow cytogenetic analysis (Castiglia et al. 2002; Solano et al. 2009). The islands/populations have both all-acrocentric $2n = 40$ standard (Filicudi and Salina) and Rb karyotypes (Alicudi, IALC: $2n = 34$; Lipari and Stromboli, ILIP: $2n = 26$; Panarea, IPAN: $2n = 35-37$; and Vulcano, IVUL: $2n = 26$). Among the latter, Panarea (IPAN) is the only population in which structural heterozygote individuals were found. The population from Central Italy belongs to an Rb race with a high number of metacentric chromosomes ($2n = 24$, Ancarano, IACR) and shares three Rb chromosomes with the ILIP race (Lipari and Stromboli) from the archipelago (fig. 1 and supplementary table S1, Supplementary Material online, show the Rb chromosomes shared by the eight focal populations).

ddRAD Sequencing, Mapping, and Genotyping

Genomic DNA was extracted from muscle tissue stored in ethanol using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA) following the manufacturer's protocol and including an RNase treatment to remove residual RNA. The DNA integrity of each sample was assessed by agarose gel electrophoresis and quantified using a Qubit v2.0 fluorometer (Life Technologies, Darmstadt, Germany). Approximately 200 ng of DNA per sample was used to construct ddRAD (Peterson et al. 2012) libraries following the modifications introduced by Franchini et al. (2014). DNA digestion was carried out using the restriction enzymes *Pst*I (rare cutter) and *Msp*I (frequent cutter). Two ddRAD libraries including 47 barcoded individuals each were constructed and size-selected from 380 to 480 bp in two different channels of a Pippin Prep system (Sage Science, Beverly, MA). The final libraries, further purified with $0.8\times$ AMPure XP beads (Beckman Coulter, Brea, CA), were single-end sequenced in two Illumina lanes of a HiSeq2500 platform with 101 cycles at the genome facility of Tufts University (TUCF Genomics, Boston, MA). A total of 242.3 million (M) raw reads were processed by the *process_radtags* script implemented in the Stacks v1.48 (Catchen et al. 2011) package. Individuals were demultiplexed, and low-quality reads were filtered out (options: -c -q). A final data set of 211.5 M reads (averaging 2.2 M reads per individual), each 96 bp in length (after removing the 5-bp barcode), was

aligned to the Mouse Genome assembly release GRCm38.p6 (NCBI accession number GCA_000001635.6) using the program bwa-mem v0.7.15 (Li and Durbin 2009) with default parameters. Variant and genotype calling were performed using the *ref_map* wrapper script of Stacks with default settings and applying a minimum read depth threshold of three reads to create a stack (-m parameter). On average, we obtained $146,743 \pm 47,621$ (SD) loci per individual with a mean coverage of 11.6 ± 4.2 (SD) reads per locus per individual after excluding five samples due to exceptionally low read or locus coverage. Unless otherwise noted, all analyses were performed on the 19 mouse autosomes only. Variant files were handled and filtered with VCFtools v0.1.15 (Danecek et al. 2011) and Plink v1.9 (Purcell et al. 2007).

Overall Population Structure and Genetic Differentiation

We applied a combination of different approaches to assess the degree of overall population structure and pairwise genetic differentiation among our samples. For the former, loci with more than 10% missing data were removed and only one marker per 1 kb was retained (1-kb thinning) to reduce the effect of nonindependence of markers due to physical linkage. For the latter, all called variable sites were retained, except for markers with more than 20% missing data in any focal population or a minor allele frequency of <0.1 . First, to investigate the possible presence of genetic clustering in the whole data set, we used the model-based algorithm implemented in Admixture v1.23 (Alexander et al. 2009). Admixture analyses were run with default parameters and allowing the number of predefined genetic clusters (K) to range from 1 to 12. Statistical support for the different number of inferred clusters was assessed using the cross-validation technique implemented in Admixture. Second, a PCA was run with the program Eigensoft v5.0.2 (Patterson et al. 2006). The *scatterplot3d* (Ligges and Mächler 2003) library run in the R v3.2.2 (R-Core-Team 2015) environment was used to plot PCA scores. Third, a phylogenetic split network was generated based on uncorrected P distances using SplitsTree v4.13.1 (Huson and Bryant 2006). The support for the splits in the network was assessed with 1,000 bootstrap replicates. Finally, pairwise genetic differentiation among populations was determined using VCFtools.

Phylogenetic Reconstruction

To reconstruct the phylogenetic relationships, we used the program SVDquartets (Chifman and Kubatko 2014) implemented in the PAUP* v4a159 package (Swofford 2001), evaluating all possible quartets and assessing robustness with 500 bootstrap replicates. SVDquartets allows the estimation of phylogenetic relationships using single-nucleotide polymorphism (SNP) data in a coalescent framework. In order to root the phylogenetic tree, a whole genome-sequenced (WGS) individual of *Mus musculus* from the Czech Republic (European Nucleotide Archive, sample accession ID: ERS957496, depth of coverage $\sim 30\times$) (Harr et al. 2016) was included for this analysis. Because we wanted to combine ddRAD and WGS reads and call genotypes at every callable

position that satisfied our criteria without the need to create read stacks across the entire genome (requirement of the program Stacks), we used the dDocent v2.2.25 pipeline (Puritz et al. 2014) with default parameters. Briefly, dDocent called Trimmomatic v0.36 (Bolger et al. 2014), bwa-mem v0.7.15 (Li and Durbin 2009), and FreeBayes v1.1.0 (Garrison and Marth 2012) to perform quality control and trimming of raw reads, align the filtered sequences to the reference genome and infer individual genotypes at polymorphic loci. We further set all individual genotypes supported by a read depth smaller than three to missing and only retained loci with a maximum of 10% missing data. Again, a 1-kb thinning procedure was applied. The final phylogenetic data set comprised 12,219 polymorphic sites.

Historical Relationships among Populations

The program TreeMix v1.12 (Pickrell and Pritchard 2012) was used to infer patterns of population splits and mixtures in the history of the house mouse populations of this study. To this end, a maximum likelihood phylogenetic tree was built from allele frequency data (these are the same data that were used for population structure analyses; see above), grouping together blocks of ten adjacent SNPs ($-k$ 10) to account for physical linkage in the implemented jackknife resampling procedure. As genomic region-dependent analyses suggested no (recent) shared ancestry between the islands populations and Ancarano from Central Italy (see Results section), this latter population was used to root the tree.

Genomic Region-Dependent Population Differentiation and Divergence

We applied Admixture and PCA approaches separately to subsets of centromeric and telomeric loci in selected populations to investigate potential linkage-dependent genetic structure. This approach was used as loci linked to metacentric centromeric regions are most likely to reflect the evolutionary history of populations harboring such chromosomes (Forster et al. 2013; Gimenez et al. 2016). For each autosomal chromosome, we created data sets including loci falling within the first (proximal loci) and last (distal loci) 10 centimorgan (cM), respectively, using the *vcfintersect* script of the *vcflib* package (<https://github.com/vcflib/vcflib>; last accessed December 2019). The Mouse Genome Database (Smith et al. 2018) was used to link recombination distance (cM) to physical distance (Mb), and thus retrieve the target genomic regions for each chromosome.

As F_{ST} -values were already calculated at each site for each chromosome, these values were analyzed and compared for the same proximal and distal 10 cM regions used in the Admixture and PCA analyses. These subsets of loci were used to test different evolutionary scenarios that could have shaped the karyotypic and genetic structures of the house mouse populations.

We calculated average pairwise sequence divergence (D_{XY}) and net nucleotide divergence (D_A) (Nei 1987) between all pairs of populations and separately for the centromeric and telomeric regions of each autosome. In contrast to F_{ST} , D_{XY} and D_A are absolute measures of population differentiation,

which require accurate estimates of the proportion of variable sites along the genome. To achieve this, we calculated site-wise read coverage for each sequenced individual using the *depth* subcommand of SAMtools (Li et al. 2009) with a minimum base quality filter of 20 and a minimum mapping quality filter of 30. Using this site-wise coverage data in combination with the genotype calls from the dDocent pipeline, we applied a custom Perl script to calculate mean pairwise sequence differences within (π) and between populations (D_{XY}) for all pairwise population comparisons. We excluded sites where more than 20% of individuals had coverage of less than three reads in any of the two populations for a given comparison or where the number of observed alleles over all individuals was larger than 2. We treated all genotypes at variant sites as missing for the calculation of pairwise differences if the coverage after input read filtering was < 3 .

Acknowledgments

The authors thank Lénia Beck for laboratory assistance. They also thank Julia Jones, Carmelo Fruciano, and Darrin Hulsey for comments and discussions. The study was funded by a FARI Research Grant no. C26I11H2MT (“La Sapienza” University of Rome, 2011). The authors acknowledge Luigi Maiorano for supporting the project. They also thank the scientific computing cluster (SCC) of the University of Konstanz.

Author Contributions

P.F., A.F.K., and A.N. devised the analytical framework and performed the bioinformatic analyses. P.F., G.A., R.C., and E.S. collected the material and conceived the study. P.F. wrote the manuscript with input from all authors.

References

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.

Berrios S, Fernandez-Donoso R, Page J, Ayarza E, Capanna E, Solano E, Castiglia R. 2018. Hexavalents in spermatocytes of Robertsonian heterozygotes between *Mus m. domesticus* 2n-26 from the Vulcano and Lipari Islands (Aeolian Archipelago, Italy). *Eur J Histochem.* 62:83–87.

Bidau CJ, Gimenez MD, Palmer CL, Searle JB. 2001. The effects of Robertsonian fusions on chiasma frequency and distribution in the house mouse (*Mus musculus domesticus*) from a hybrid zone in northern Scotland. *Heredity* 87(3):305–313.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

Bush GL, Case SM, Wilson AC, Patton JL. 1977. Rapid speciation and chromosomal evolution in mammals—(evolutionary rates effective population size vertebrates cytogenetic isolating mechanisms social structuring). *Proc Natl Acad Sci U S A.* 74(9):3942–3946.

Capanna E, Redi CA. 1995. Whole-arm reciprocal translocation (WART) between Robertsonian chromosomes—finding of a Robertsonian

heterozygous mouse with karyotype derived through WARTs. *Chromosome Res.* 3(2):135–137.

Castiglia R, Annesi F, Capanna E. 2002. Contact zones between chromosomal races of *Mus musculus domesticus*. 3. Molecular and chromosomal evidence of restricted gene flow between the CD race ($2n = 22$) and the ACR race ($2n = 24$). *Heredity* 89(3):219–224.

Castiglia R, Capanna E. 2002. Chiasma repatterning across a chromosomal hybrid zone between chromosomal races of *Mus musculus domesticus*. *Genetica* 114(1):35–40.

Castiglia R, Capanna E, Bezerra AMR, Bizzoco D, Zambigli E, Solano E. 2015. New metacentric populations and phylogenetic hypotheses involving whole-arm reciprocal translocation in *Mus musculus domesticus* from Sicily. *Cytogenet Genome Res.* 146(3):230–237.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3 (Bethesda)* 1:171–182.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134(4):1289–1303.

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30(23):3317–3324.

Coyne JA, Orr HA. 2004. Speciation. Sunderland: Sinauer Associates.

Cucchi T, Vigne JD, Auffray JC. 2005. First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. *Biol J Linn Soc.* 84(3):429–445.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.

Dumas D, Britton-Davidian J. 2002. Chromosomal rearrangements and evolution of recombination: comparison of chiasma distribution patterns in standard and Robertsonian populations of the house mouse. *Genetics* 162(3):1355–1366.

Evans JA, de von Flindt R, Greenberg C, Ramsay S, Hamerton JL. 1982. A cytogenetic survey of 14,069 newborn-infants. 4. Further follow-up on the children with sex-chromosome anomalies. *Birth Defects Orig Artic Ser* 18(4):169–184.

Faria R, Navarro A. 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol Evol.* 25(11):660–669.

Forster DW, Jones EP, Johannesdottir F, Gabriel SI, Gimenez MD, Panithanarak T, Hauffe HC, Searle JB. 2016. Genetic differentiation within and away from the chromosomal rearrangements characterising hybridising chromosomal races of the western house mouse (*Mus musculus domesticus*). *Chromosome Res.* 24(2):271–280.

Forster DW, Mathias ML, Britton-Davidian J, Searle JB. 2013. Origin of the chromosomal radiation of Madeiran house mice: a microsatellite analysis of metacentric chromosomes. *Heredity* 110(4):380–388.

Franchini P, Colangelo P, Meyer A, Fruciano C. 2016. Chromosomal rearrangements, phenotypic variation and modularity: a case study from a contact zone between house mouse Robertsonian races in Central Italy. *Ecol Evol.* 6(5):1353–1362.

Franchini P, Colangelo P, Solano E, Capanna E, Verheyen E, Castiglia R. 2010. Reduced gene flow at pericentromeric loci in a hybrid zone involving chromosomal races of the house mouse *Mus musculus domesticus*. *Evolution* 64:2020–2032.

Franchini P, Fruciano C, Spreitzer ML, Jones JC, Elmer KR, Henning F, Meyer A. 2014. Genomic architecture of ecologically divergent body shape in a pair of sympatric crater lake cichlid fishes. *Mol Ecol.* 23(7):1828–1845.

Gabriel SI, Johannesdottir F, Jones EP, Searle JB. 2010. Colonization, mouse-style. *BMC Biol.* 8(1):131.

Garagna S, Marziliano N, Zuccotti M, Searle JB, Capanna E, Redi CA. 2001. Pericentromeric organization at the fusion point of mouse Robertsonian translocation chromosomes. *Proc Natl Acad Sci U S A.* 98(1):171–175.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

- Gazave E, Catalan J, Ramalhinho MD, Mathias MD, Nunes AC, Dumas D, Britton-Davidian J, Auffray JC. 2003. The non-random occurrence of Robertsonian fusion in the house mouse. *Genet Res.* 81(1):33–42.
- Gimenez MD, Panithanarak T, Hauffe HC, Searle JB. 2016. Empirical demonstration of hybrid chromosomal races in house mice. *Evolution* 70(7):1651–1658.
- Harr B, Karakoc E, Neme R, Teschke M, Pfeifle C, Pezer Ž, Babiker H, Linnenbrink M, Montero I, Scavetta R, et al. 2016. Genomic resources for wild populations of the house mouse, *Mus musculus* and its close relative *Mus spretus*. *Sci Data* 3:160075.
- Hauffe HC, Gimenez MD, Vega R, White TA, Searle JB. 2011. Properties of a hybrid zone between highly distinct chromosomal races of the house mouse (*Mus musculus domesticus*) in northern Italy, and comparisons with other hybrid zones. *Cytogenet Genome Res.* 134(3):191–199. 2011.
- Hauffe HC, Pialek J. 1997. Evolution of the chromosomal races of *Mus musculus domesticus* in the Rhaetian Alps: the roles of whole-arm reciprocal translocation and zonal raiation. *Biol J Linn Soc.* 62(2):255–278.
- Henikoff S, Ahmad K, Malik HS. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293(5532):1098–1102.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 23(2):254–267.
- Irwin DE, Mila B, Toews DPL, Brelsford A, Kenyon HL, Porter AN, Grossen C, Delmore KE, Alcaide M, Irwin JH. 2018. A comparison of genomic islands of differentiation across three young avian species pairs. *Mol Ecol.* 27(23):4839–4855.
- King M. 1993. Species evolution: the role of chromosome change. Cambridge: Cambridge University Press.
- Kirkpatrick M. 2017. The evolution of genome structure by natural and sexual selection. *J Hered.* 108(1):3–11.
- Leache AD, Banbury BL, Linkem CW, de Oca A. 2016. Phylogenomics of a rapid radiation: is chromosomal evolution linked to increased diversification in North American spiny lizards *BMC Evol Biol.* 16:63.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Li H, Ralph P. 2019. Local PCA shows how the effect of population structure differs along the genome. *Genetics* 211(1):289–304.
- Ligges U, Mächler M. 2003. Scatterplot3d—an R package for visualizing multivariate data. *J Stat Softw.* 8(11):1–20.
- Lindholm AK, Dyer KA, Firman RC, Fishman L, Forstmeier W, Holman L, Johannesson H, Knief U, Kokko H, Larracuenta AM, et al. 2016. The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol Evol.* 31(4):315–326.
- Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
- Pagacova E, Cernohorska H, Kubickova S, Vahala J, Rubes J. 2011. Centric fusion polymorphism in captive animals of family Bovidae. *Conserv Genet.* 12(1):71–77.
- Panithanarak T, Hauffe HC, Dallas JF, Glover A, Ward RG, Searle JB. 2004. Linkage-dependent gene flow in a house mouse chromosomal hybrid zone. *Evolution* 58(1):184–192.
- Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2(12):e190.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7(5):e37135.
- Pialek J, Hauffe HC, Searle JB. 2005. Chromosomal variation in the house mouse. *Biol J Linn Soc.* 84(3):535–563.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8(11):e1002967.
- Potter S, Bragg JG, Blom MPK, Deakin JE, Kirkpatrick M, Eldridge MDB, Moritz C. 2017. Chromosomal speciation in the genomics era: disentangling phylogenetic evolution of rock-wallabies. *Front Genet.* 8: 10.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575.
- Puritz JB, Hollenbeck CM, Gold JR. 2014. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* 2:e431.
- R-Core-Team. 2015. A language and environment for statistical computing. Vienna (Austria). Available from: <http://www.R-project.org/>.
- Redi CA, Garagna S, Zuccotti M. 1990. Robertsonian chromosome formation and fixation—the genomic scenario. *Biol J Linn Soc.* 41(1-3):235–255.
- Ribagorda M, Berrios S, Solano E, Ayarza E, Martin-Ruiz M, Gil-Fernandez A, Parra MT, Viera A, Rufas JS, Capanna E, et al. 2019. Meiotic behavior of a complex hexavalent in heterozygous mice for Robertsonian translocations: insights for synapsis dynamics. *Chromosoma* 128(2):149–163.
- Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends Ecol Evol.* 16(7):351–358.
- Riginos C, Nachman MW. 1999. The origin of a Robertsonian chromosomal translocation in house mice inferred from linked microsatellite markers. *Mol Biol Evol.* 16(12):1763–1773.
- Rosin LF, Mellone BG. 2017. Centromeres drive a hard bargain. *Trends Genet.* 33(2):101–117.
- Searle JB. 1993. Chromosomal hybrid zones in eutherian mammals. In: Harrison RG, editor. *Hybrid zones and the evolutionary process*. Oxford: Oxford University Press. p. 309–353.
- Smith CL, Blake JA, Kadin JA, Richardson JE, Bult CJ; Mouse Genome Database Group 2018. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Res.* 46(D1):D836–D842.
- Solano E, Castiglia R, Capanna E. 2009. Chromosomal evolution of the house mouse, *Mus musculus domesticus*, in the Aeolian Archipelago (Sicily, Italy). *Biol J Linn Soc.* 96(1):194–202.
- Solano E, Castiglia R, Corti M. 2007. A new chromosomal race of the house mouse, *Mus musculus domesticus*, in the Vulcano Island-Aeolian Archipelago. Italy. *Hereditas* 144(3):75–77.
- Solano E, Franchini P, Colangelo P, Capanna E, Castiglia R. 2013. Multiple origins of the western European house mouse in the Aeolian Archipelago: clues from mtDNA and chromosomes. *Biol Invasions* 15(4):729–739.
- Song JP, Li X, Sun L, Xu SQ, Liu N, Yao YY, Liu Z, Wang WP, Rong H, Wang B. 2016. A family with Robertsonian translocation: a potential mechanism of speciation in humans. *Mol Cytogenet.* 9:48.
- Swofford D. 2001. Phylogenetic analysis using parsimony (*and other methods), version 4.0a142. Sunderland (MA): Sinauer.
- Talbert PB, Henikoff S. 2010. Centromeres convert but don't cross. *PLoS Biol.* 8(3):e1000326.
- White M. 1973. Animal cytology and evolution. Cambridge: Cambridge University Press.
- White TA, Bordewich M, Searle JB. 2010. A network approach to study karyotypic evolution: the chromosomal races of the common shrew (*Sorex araneus*) and house mouse (*Mus musculus*) as model systems. *Syst Biol.* 59(3):262–276.