AFLP genome scan in the black rat (*Rattus rattus*) from Madagascar: detecting genetic markers undergoing plague-mediated selection

C. TOLLENAERE,* J.-M. DUPLANTIER,* L. RAHALISON,† M. RANJALAHY† and C. BROUAT*

*IRD UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Campus International de Baillarguet, CS 30016, 34988 Montferrier-sur-Lez, Cedex, France, †Institut Pasteur de Madagascar (IPM), Unité Peste, BP1274, Ambatofotsikely, 101 Antananarivo, Madagascar

Abstract

The black rat (*Rattus rattus*) is the main reservoir of plague (*Yersinia pestis* infection) in Madagascar's rural zones. Black rats are highly resistant to plague within the plague focus (central highland), whereas they are susceptible where the disease is absent (low altitude zone). To better understand plague wildlife circulation and host evolution in response to a highly virulent pathogen, we attempted to determine genetic markers associated with plague resistance in this species. To this purpose, we combined a population genomics approach and an association study, both performed on 249 AFLP markers, in Malagasy *R. rattus*. Simulated distributions of genetic differentiation were compared to observed data in four independent pairs, each consisting of one population from the plague focus and one from the plague-free zone. We found 22 loci (9% of 249) with higher differentiation in at least two independent population pairs or with combining *P*-values over the four pairs significant. Among the 22 outlier loci, 16 presented significant association with plague zone (plague focus *vs.* plague-free zone). Population genetic structure inferred from outlier loci was structured by plague zone, whereas the neutral loci dataset revealed structure by geography (eastern *vs.* western populations). A phenotype association study revealed that two of the 22 loci were significantly associated with differentiation between dying and surviving rats following experimental plague challenge. The 22 outlier loci identified in this study may undergo plague selective pressure either directly or more probably indirectly due to hitchhiking with selected loci.

Keywords: adaptation genetics, association study, disease resistance, pathogen-mediated selection, population genomics, *Yersinia pestis*

Received 30 September 2009; revision received 1 March 2010; revision accepted 7 March 2010

Introduction

Population genomics (Black *et al.* 2001; Luikart *et al.* 2003) offers the opportunity to address the question of adaptation genetics by separating locus-specific effects from genome-wide demographic effects (Cavalli-Sforza 1966). Loci under natural selection are expected to present higher genetic differentiation than neutral loci between populations suffering different selective envi-ronments (Beaumont 2005). Population genome scans were recently performed in various wild species (see Nosil *et al.* 2009 for a review) to detect genetic markers associated with selective pressure (Bonin *et al.* 2006; Williams & Oleksiak 2008), with phenotypic differences (Herrera & Bazaga 2008), or with both adaptive phenotypic change and selective pressure (for example, ecotypes of *Coregonus* varying in depth habitat, Campbell & Bernatchez 2004). In some cases, the validity of the population genomics approach was confirmed by the co-location of identified loci with quantitative trait loci (QTL) (Campbell & Bernatchez 2004; Via & West 2008)

Correspondence: Carine Brouat, Fax: +33 (0) 4 99 62 33 45;
E-mail: brouat@mpl.ird.fr

or by individual-based phenotype association studies (Storz 2005). Amplified Fragment Length Polymorphism (AFLP) markers (Vos *et al.* 1995) are generally used for such genomic studies of non-model species because of the possibility they offer for developing a large number of loci without any knowledge of the species genome and with limited cost (Meudt & Clarke 2007).

To our knowledge, AFLP genome scans have not been applied to detect genetic markers associated with pathogen resistance in host populations. Parasites can exert strong selective pressure on their hosts, resulting in extremely rapid evolutionary changes (Altizer *et al.* 2003). Consequently, genes involved in immune defence have been shown to be strongly affected by natural selection (Nielsen *et al.* 2005; Piertney & Oliver 2006; Lazzaro 2008). Investigating the genetic basis of host resistance/susceptibility has inherent interest for our understanding of the role of parasites in host evolution as well as important health applications in the context of emerging and re-emerging diseases. Host resistance to infectious diseases is usually considered a highly polygenic trait (Sorci *et al.* 1997; Hill 2001). However, QTL studies of disease resistance in animals have usually identified only a few loci (Wilfert & Schmid-Hempel 2008). Although most of the studies dealing with host-pathogen interaction genetics in natural vertebrate populations focused on the major histocompatibility complex (MHC), it is now recognized that variation in the MHC region does not sufficiently explain wildlife immunogenetics (Acevedo-Whitehouse & Cunningham 2006). In the context of pathogen resistance, AFLP genome scans thus seem a relevant approach to identification of genomic regions potentially undergoing natural selection at the genome scale, with no need of *a priori* knowledge about host defence mechanisms or the genes involved.

Plague (*Yersinia pestis* infection) is among the most virulent of known pathogens for humans and susceptible rodents (Stenseth *et al.* 2008); its epidemics may consequently exert very strong selective pressure on its hosts (see for example Thomas *et al.* 1988; Hanson *et al.* 2007). The disease was introduced into Madagascar in 1898. Areas of the central highlands of this country with altitudes higher than 800 meters have represented a plague focus since the 1920s, with hundreds of human cases reported each year and circulation within rodent populations, whereas the low-altitude zone of Madagascar presents no evidence of plague circulation in rural areas (Brygoo 1966; Migliani *et al.* 2006). The main reservoir of the disease in Madagascar is the black rat, *Rattus rattus* (Brygoo 1966; Duplantier *et al.* 2003), a species usually considered to be highly susceptible to plague (Dennis *et al.* 1999). However, Rahalison *et al.* (2003) and Tollenaere *et al.* (pers. obs.) showed that *R. rattus*

populations from the plague focus (central highlands) are much more resistant (1000-fold difference in the LD_{50}) than black rat populations from plague-free zones (no *Y. pestis* circulation). This resistance is transmitted to laboratory-born descendants (Rahalison *et al.* 2003), revealing its genetic basis. Plague resistance variability in *R. rattus* populations could potentially have high impact on plague wildlife circulation, as the proportion of individuals susceptible/resistant to plague in a population determines the role the population can play in transmission/maintenance of the disease (Gage & Kosoy 2005). Besides being of theoretical interest with respect to the question of how parasites shape genetic variation of their hosts, identification of genetic markers associated with plague resistance in *R. rattus* would thus also be a step towards a better understanding of plague circulation in wild rat populations of the Malagasy focus.

Although a plague resistance locus was identified within the MHC region in laboratory mice strains (Turner *et al.* 2008), it does not necessarily follow that the same locus would be implicated in other contexts because genetic determinism may differ even between different host lines (Wilfert & Schmid-Hempel 2008). Moreover, detailed mechanisms of plague resistance remain poorly understood (Turner *et al.* 2008) and have never been studied in Malagasy *R. rattus*, making hazardous the definition of candidate genes (but see Tollenaere *et al.* 2008; on CCR5 candidate gene). On the other hand, this context appears highly relevant for a population genomics approach (Storz & Hoekstra 2007; Ellegren & Sheldon 2008) as (i) the selective pressure (plague disease) is identified and probably strong (high mortality rate in rodent populations); (ii) adaptive phenotypic change (plague resistance) is characterized; (iii) the studied species diverged recently from the model species *R. norvegicus* (about 2.9 Myr, Robins *et al.* 2008) and as such a good genomic resource is available to identify AFLP markers (sequences of interesting AFLP fragments can be localized by blasting on the *R. norvegicus* genome).

The goal of this study was to identify markers potentially associated with plague resistance in Malagasy *R. rattus* populations using an AFLP genome scan. To this end, we sampled independent natural populations that inhabit environments with contrasting or similar selective pressures (see also Williams & Oleksiak 2008): four independent pairs of populations, each consisting of one population located in the plague focus (central highlands of Madagascar) and one population located in the plague-free zone (low altitude zone of the island) were analyzed. We also investigated associations between AFLP markers and phenotypic variation within the plague focus: the genetics of individuals that

survived or died following experimental plague challenge was compared.

Materials and methods

Sampling

Black rats were sampled between April 2006 and January 2009 in eight sites on Madagascar (Fig. 1): four in the central highlands (plague focus, Ambohitseheno, AIT, 1415 m, Antahobe, ANT, 1200 m, Inanantonana,

INA, 1268 m and Morarano-Gare, MOR, 906 m) and four in the low altitude zone (plague-free zone, Ankotrofotsy, ANK, 85 m, Brickaville, BRI, 12 m, Miandrivazo, MIA, 72 m, and Tsarasambo, TSO, 15 m). This resulted in four independent pairs populations, each consisting of one population from the plague focus and one population from the plague-free zone (Fig. 1), with two pairs in the east of the island: E1 (AIT-TSO, 138.1 km distant from each other) and E2 (BRI-MOR, 85.0 km apart), and two pairs in the west: W1 (ANK-ANT, 122.8 km apart) and W2 (INA-MIA, 123.6 km apart). The geographic distribution of population pairs across the east and west sides of Madagascar was chosen to disentangle the correlated effects of plague presence/absence and other potential selective factors (see details in the Discussion), as low altitude east and west Malagasy zones present different environmental conditions (Jury 2003).

At least 25 rats were trapped in each site using wire-meshed and Sherman traps set in houses and outdoors. Caught animals were euthanized by cervical dislocation or bred for other experiments. Tissues (ear, tail or finger) were kept in 95° ethanol for genetic analyses.

Laboratory procedures

Total DNA was extracted using the DNeasy® Tissue Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions for rodent tails and animal tissues. DNA was eluted in 100 μL of elution buffer. DNA concentration was estimated by a fluorimetric quantification of dsDNA using the Picogreen Kit (Molecular Probes, Invitrogen, UK). Restriction enzymes *Eco*RI and *Taq*I were used following the protocol described in Ajmone-Marsan *et al.* (1997). Briefly, 350 ng of DNA was first incubated with five units *Taq*I (Promega, Madison, USA) for 2 h at 65 °C. The product was then incubated with five units *Eco*RI (Promega) for 1 h at 37 °C. Finally, the solution was incubated for 3 h at 37 °C with one unit DNA ligase (Promega) and adapters for *Taq*I (5′-CGGTCAGGACTCAT-3′ and 5′-GACGATGAGTCC TGAC-3′, 50 pmol) and *Eco*RI (5′-CTCGTAGACTGCGT ACC-3′ and 5′-AATTGGTACGCAGTCTAC-3′, 5 pmol) restriction sites. All reactions were performed in the recommended enzyme buffers with bovine serum albumin (BSA) according to the manufacturer's instructions. DNA was diluted 1:10 in water after digestion and ligation. Pre-selective PCR was performed using primers E-A (5′-GACTGCGTACCAATTCA-3′) and T-A (5′-GATGAGTCCTGACCGAA-3′). Amplified DNA was diluted 1:20 before performing selective PCR using 10 different primer pairs (Table 1). Amplified fragments were detected using an automated Prism 3130XL sequencer (Applied Biosystems).
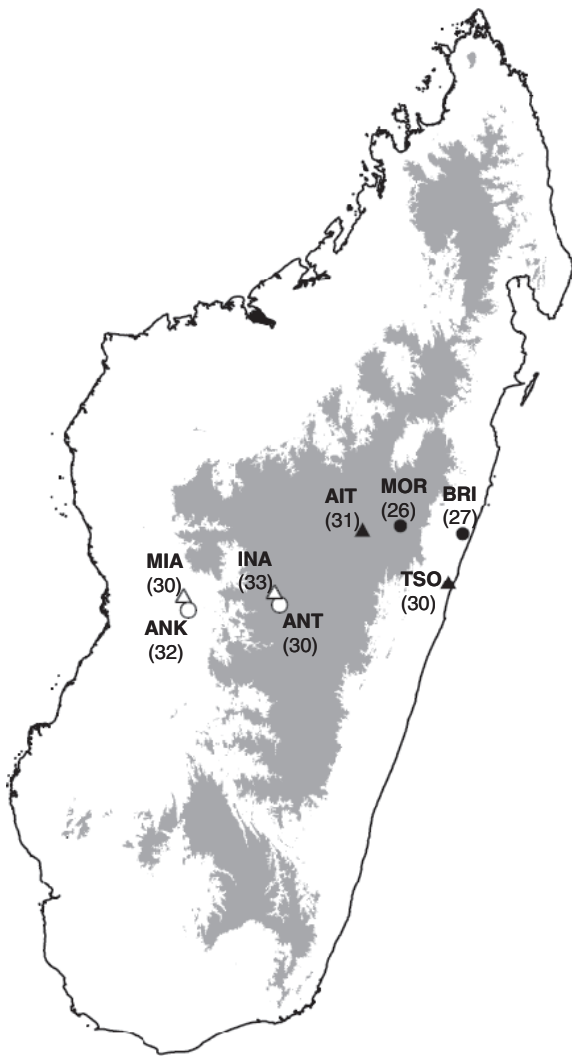

Fig. 1 Sampling map indicating the location of the eight *R. rattus* populations analyzed. Populations from the same pair are indicated with the same symbol: black triangle (E1), black circle (E2), white circle (W1) and white triangle (W2). The sample size is shown in brackets for each population. The central highlands, where the altitude is higher than 800 meters, is coloured in grey. This zone corresponds to the plague focus (Brygoo 1966; Duplantier *et al.* 2005).

Table 1 Combinations of primers used for selective amplification in the AFLP procedure. Each primer *Eco*RI begins with 5′-GACTGCGTACCAATTC-3′ and each primer *Taq*I with 5′-GAT-GAGTCCTGACCGA-3′, followed by the three selective bases indicated in this table.

	Combination	Primer *Eco*RI	Primer *Taq*I	Number of markers
1	A	E-AAG	T-AAG	43
2	B	E-AAG	T-AGA	23
3	C	E-AAG	T-ATC	21
4	D	E-AAG	T-ATG	27
5	J	E-ACT	T-AAG	35
6	N	E-ACT	T-AGT	26
7	Q	E-ATA	T-AAG	21
8	U	E-ATA	T-ATG	22
9	W	E-AAG	T-AGT	22
10	Y	E-ATA	T-AGT	9

AFLP scoring

Electrophoresis reading was performed using GeneMapper 3.7 software (Applied Biosystems). The choice and scoring of markers is a critical part of the AFLP procedure (Bonin *et al.* 2005). As genotyping errors are more frequent for small fragments (Meyer *et al.* 2009), peaks were scored between 60 and 520 bp. We then chose loci using two criteria: polymorphism (band presence or band absence in at least two individuals) and clarity of reading (clear threshold between presence/absence and no peak superposition). To ensure quality of the data (Pompanon *et al.* 2005), 16 individuals (6.7% of total individuals) were genotyped twice from the same sample following the extraction step. The error rate was estimated to be 0.63% (25 errors out of 3984 comparisons).

Outlier detection simulations

For each pair of populations analyzed, we estimated genetic differentiation (F_{ST}) using AFLP-surv 1.0 software (Vekemans 2002). Exact tests for differentiation between population pairs were performed on marker frequencies using TFPGA v. 1.3 software (Mark Miller 1997, available at http://www.marksgeneticsoftware.net/).

We used two programs allowing the detection of loci that showed high differentiation compared to the whole dataset (Beaumont 2005). Each approach was applied to the four independent comparisons of 'plague-focus' and 'plague-free' populations: E1 (AIT/TSO), W1 (ANK/ANT), E2 (BRI/MOR) and W2 (INA/MIA). This sampling design, comparing independent population pairs, was originally recommended by Tsakas & Krimbas (1976) (see also Vitalis *et al.* 2001). We first used the

program *DetSelD* (available upon request to the author: vitalis@supagro.inra.fr), a new version of the *DetSel* software (Vitalis *et al.* 2001, 2003) for analysing dominant markers such as AFLP (Midamegbe *et al.* submitted). *DetSelD* performs coalescent simulations using a pure divergence model in which an ancestral population suffers a bottleneck and then splits into two isolated populations. The null distribution was obtained using 20 combinations of nuisance parameters values: we allowed variations in the ancestral population size ($Ne = 10^3$, 10^4, 10^5 and 10^6 individuals) and divergence time ($t = 50$, 100, 500, 1000 and 5000 generations); the bottleneck duration was set to one generation and the population size after bottleneck was 1/10 of the ancestral population size. For each pairwise comparison and each nuisance parameter set, 100 000 simulations were performed. Second, we used the *DFdist* program (available at http://www.rubic.rdg.ac.uk/~mab/stuff/), an extension for dominant markers of the original *Fdist* program (Beaumont & Nichols 1996). *DFdist* first estimates genetic differentiation (F_{ST}) for each locus using Zhivotovsky's (1999) Bayesian method to estimate allelic frequencies. This empirical distribution allows the estimation of the global estimation of 'neutral' population differentiation as the trimmed (30% highest and lowest F_{ST} values removed) mean F_{ST} (recommended by Beaumont & Balding 2004). This global F_{ST} is then used to perform 50 000 simulations following an island model (2 demes). Both programs (*DetSelD* and *DFdist*) were used with a threshold of 5%.

Classification and analyses of outliers

P-values obtained for the four independent comparisons were combined following Fisher's method (Fisher 1948). [Some *DFdist* P-values were equal to zero and we conservatively changed them to 0.0001 to minimize their effects on Fisher's P-values]. We considered as outlier loci those that, for at least one detection program (*DFdist* or *DetselD* approach), either (i) were detected as outliers in at least two independent comparisons at the threshold 5%, i.e. 'repeated outliers' or (ii) had Fisher's combined P-values lower than 1%, i.e. 'combined outliers'. Indeed, the repetition of outlier detection in various comparison is widely used (Storz 2005; see for example Bonin *et al.* 2006), as replicated divergence across population pairs is unlikely to arise via nonselective factors such as type I error, genetic drift, or mutation rate variation (Nosil *et al.* 2009). Whereas various studies considered outliers repeated in non-independent comparisons (for example Nosil *et al.* 2008), we only considered repetition in independent comparison. Moreover, combining P-values is another method recently used (Meyer *et al.* 2009); this method can

detect some loci presenting *P*-values close to the threshold in various comparisons which appear to be good candidates but would not be detected by the other method.

For each outlier locus, we tested for association between the presence/absence of the peak and plague zone (plague focus *vs.* plague-free zone). To this end, we performed Fisher's exact tests using R software (Ihaka & Gentleman 1996). This method is favoured because it allows testing for independence between two binary variables in cases of low sample size, which may occur here for loci with extreme peak frequency. Correction for multiple testing was performed following the false discovery rate approach (FDR, which is the proportion of false positives among the tests found to be significant, Benjamini & Hochberg 1995). To this purpose, the software *Q*-value (available at http://genomics.princeton.edu/storeylab/qvalue/) was used, fixing the tuning parameter λ to 0 (Benjamini & Hochberg 1995).

The proportion of outlier loci in each primer pair combination were tested for homogeneity using χ^2 test, as various studies reported non-random repartition of outlier loci among primer combinations (Campbell & Bernatchez 2004; Herrera & Bazaga 2008).

Genetic structure compared between neutral and outlier datasets

Genetic structures were compared between the neutral dataset, the outlier dataset and a subset of the outlier dataset comprising only the outlier loci with significant association with the zone (plague focus *vs.* plague-free zone). For each set of markers, analyses of Molecular Variance (AMOVA) were conducted using Arlequin v. 3.01 (Excoffier *et al.* 2005) on populations grouped into the two zones (plague focus *vs.* plague-free zone), or, alternatively, grouped into the longitude location (east *vs.* west, Fig. 1). We also generated 1000 bootstrapped Nei's distances using AFLP-surv software (Vekemans 2002). Neighbor-joining trees were constructed with Neighbor and Consense programs within the package Phylip v. 3.2. (http://evolution.gs.washington.edu/phylip.html).

Genotyping and analysis of plague-challenged individuals

Two datasets consisting of plague-challenged individuals were also analyzed for AFLP markers using the same genotyping protocol. The plague challenge protocol is detailed in Tollenaere *et al.* (2008). Briefly, the same strain of *Yersinia pestis* was injected subcutaneously into wild-caught animals from villages located in the plague focus (and thus expected to present resistant individuals), close to the INA locality (Fig. 1). Only rats that did not present plague-specific antibodies (supposed plague naïve) were plague-challenged. The dose of bacteria injected was either 100 cfu (colony forming units), which kills all animals from low-altitude zones, or 10^5 cfu, the LD_{50} (lethal dose for 50% of individuals) of rats from plague focus areas (Rahalison *et al.* 2003). Animals were followed for 18 days post-infection. Dead animals were tested for plague-specific antigens to assess that plague was the cause of death; surviving animals were tested for plague antibodies to verify the infection success.

The 'high dose' dataset consisted of 54 animals inoculated with 10^5 cfu of *Yersinia pestis*; 27 of these animals survived and 27 died from plague (Tollenaere *et al.* 2008). The 'low dose' dataset included 49 animals inoculated with 100 cfu of *Yersinia pestis*; of these, 35 animals survived and 14 died (Tollenaere *et al.* 2008). Of the 249 AFLP markers, 226 were polymorphic within the 'high dose' dataset and 223 within the 'low dose' dataset. For each dataset and for each AFLP marker, Fisher's exact tests were performed to test for association between the binary variable 'plague challenge issue' (dying or surviving) and the presence/absence of the band, using the R software (Ihaka & Gentleman 1996). Correction for multiple testing was performed following FDR approach, as previously described for association with the 'zone'. Relative risks (RR) were calculated following Haldane (1956). Fisher's method (Fisher 1948) was used to combine *P*-values over the two experiments.

Results

A total of 239 individuals (29.9 ± 2.4 individuals by population, see details in Fig. 1) were genotyped for 249 polymorphic loci (24.9 ± 9.1 loci by primer pair, see details in Table 1). All individuals exhibited unique AFLP profiles. AFLP loci are designated by the primer pair code (see Table 1) and the fragment length.

Outlier detection

Global F_{ST} used for *DFdist* simulations in each population pair ranged from 0.0147 to 0.0664 (Table 2). For the E2 comparison, the trimmed mean F_{ST} estimated from data was slightly negative ($F_{ST} = -0.00033$). As F_{ST} cannot be zero in the *DFdist* program, we used 0.005 as conservative global F_{ST} (see also Miller *et al.* 2007). For each comparison of plague-focus *vs.* plague-free population, we detected 14.8 ± 3.2 loci at the 5% level (and 3.8 ± 1.3 at the 1% level) using *DetSelD* and 15.3 ± 5.3 loci (and 5.3 ± 2.1 at the 1% level) using *DFdist* (Fig. 2).

Table 2 Genetic differentiation (F_{ST}) found in the four population pairs considered in this study from the whole dataset (249 markers)

Population pair name	Plague focus population	Plague-free zone population	Location	F_{ST}	Trimmed F_{ST} used for *DFdist*
E1	AIT	TSO	East	0.0281***	0.008
W1	ANT	ANK	West	0.0664***	0.048
E2	MOR	BRI	East	0.0147[NS]	**0.005**
W2	INA	MIA	West	0.0238**	0.005

Significance of F_{ST} is shown: NS non-significant, **$P < 0.01$ and ***$P < 0.0001$. The trimmed mean F_{ST} estimated by *DFdist* (removing the 30% extreme F_{ST} loci) is indicated, except for the population pair E2, for which *DFdist* estimated a slightly negative F_{ST} and the value 0.005 was arbitrarily chosen for simulations.

Eleven loci were detected in at least two independent comparisons using *DetSelD*, whereas seven were found using *DFdist* (three loci common to both programs, Table 3). Combining *DetSelD* P-values, 16 loci had significant P-values (<0.01) whereas for *DFdist*, there were ten (eight loci common to both programs, Table 3). By pooling the results of the different detection methods, we obtained a set of 22 outlier loci (Table 3). Three of the 22 outliers, A121, D466 and Y222, were detected in the four methods used (each program and 'repeated' or 'combining' method). Nine of the 22 loci were detected both in one population pair from the eastern populations (E1 or E2) and one pair from the western populations (W1 or W2) by at least one of the software methods (Table 3).

Outlier analysis

After correction for multiple testing, association with the variable 'zone' (plague focus or plague-free zone) was found (q-value < 0.05) for 16 out of 22 loci (Table 3). Among the three loci detected as outliers using the four detection methods, two had significant association with the zone (A121 and Y222, Table 3).

The proportion of outlier loci in the ten primer combinations varied between 0 (combination N) and 18.5% (combination D) (mean: 8.8 ± 5.2%). No significant difference in the proportion of outlier loci was found between the ten combinations ($\chi^2 = 7.30$, $P = 0.60$).

Comparison of genetic structure between neutral and outlier datasets

We compared genetic structure for the neutral dataset (227 loci), for the outlier dataset (22 loci) and for the outlier subdataset with significant association with zone (16 loci). AMOVAS (Table 4) revealed that the majority of genetic variance could be attributed to within-population variability whatever the dataset considered. The distinction between zones (plague focus *vs.* plague-free zone) was significant for outlier datasets only. The proportion of genetic variance attributed to the zone was 5.8% in the dataset of 22 outlier loci and 8.1% in the subdataset of 16 outlier loci. On the other hand, the distinction between longitude locations (east *vs.* west) was significant for neutral loci only (explaining 1.5% of the observed variation).

A neighbour-joining tree for the neutral dataset (227 loci) was constructed and grouped populations according to geography (Fig. 2a). It differed from the population tree obtained with outlier loci (22 loci) and from that obtained with the subset of outlier loci (16 loci), in which populations clustered according to their locations inside or outside the plague focus (Fig. 2b and c).

Analysis of plague-challenged individuals

No AFLP locus was associated with plague challenge after correction for multiple testing. However, two outlier AFLP markers for the 'high dose' dataset, B408 ($P = 0.014$) and D466 ($P = 0.005$), had P-values lower than 0.05 (Table 5). No outlier locus for the 'low dose' dataset gave P-values lower than 0.05. However, for both markers, the difference observed for the first dataset had the same tendency in the second (peak presence was associated with death for B408, whereas the opposite was found for D466) and Fisher's combined P-values were lower than 0.05 (Table 5).

One non-outlier locus was associated ($P < 0.05$) with the outcome of plague challenge in the 'high dose' dataset (B167, $P = 0.002$) and six in the 'low dose' (J220, $P = 0.019$; U434, $P = 0.021$; W162, $P = 0.025$; Y163, $P = 0.042$; U305, $P = 0.044$; N221, $P = 0.048$). Two markers, B167 ($P = 0.007$) and B172 ($P = 0.045$), had significant combined p-values. We noticed no evidence of selection on locus B167 through the population genomics approach. On the other hand, the locus B172 was an outlier in the population pair E1 when using *DetSelD* program (the combined P-value over the four pairs using *DetSelD* was 0.032).
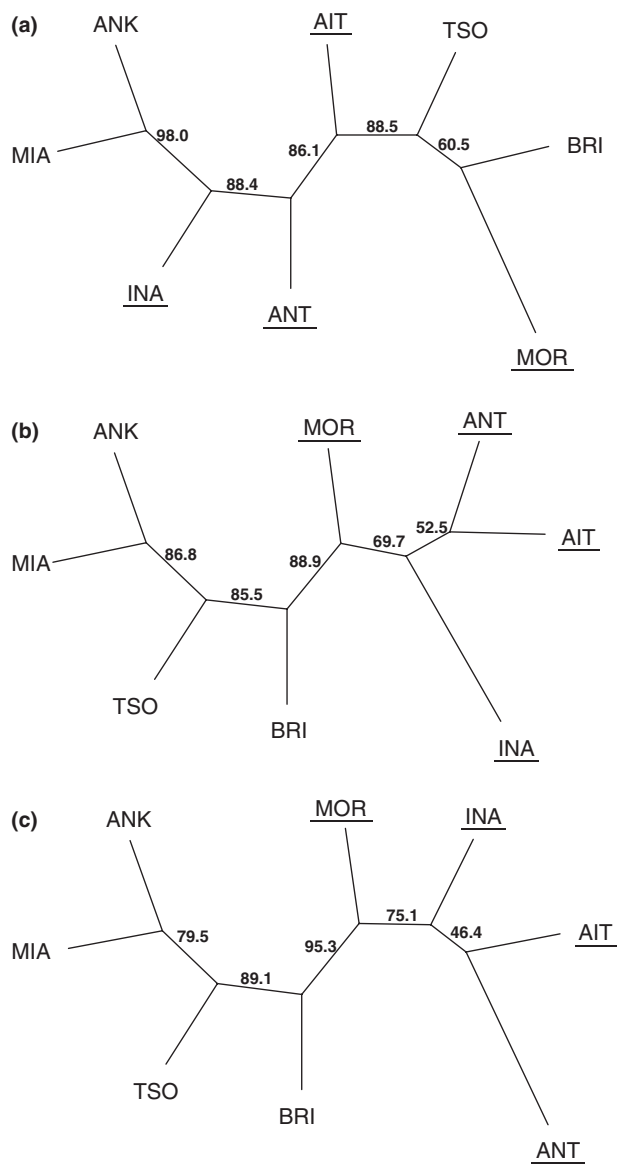
Fig. 2 Neighbour joining population trees obtained for the neutral dataset (227 loci, 2a), for the outlier dataset (22 loci, 2b) and for the subset of outlier loci for which significant association was found with the zone (16 loci, 2c). Each population is indicated by its code. Populations from plague focus are underlined. Percentages over 1000 bootstrap are indicated for each node.

Discussion

Zoonotic diseases such as plague cannot be eradicated though vaccination measures; consequently, an understanding of the evolutionary ecology of such wildlife diseases appears essential (Daszak *et al.* 2000; Stenseth *et al.* 2008). In this study, we investigated plague resistance genetics in Malagasy natural populations of its reservoir, *Rattus rattus*, using a population genomics

approach combined with a phenotype association study.

The genetic markers considered in this study are 249 AFLP loci. Because AFLP loci are anonymous genomic markers likely to be located in non-coding DNA, we did not expect to find loci under direct natural selection, but rather loci undergoing hitchhiking with selected loci (Schlotterer 2003). Hitchhiking may extend a considerable distance from the selected locus (Charlesworth *et al.* 1997; Via & West 2008). Its intensity depends on the ratio of the strength of selection and the recombination rate (Andolfatto 2001).

The number of markers analysed in this study was quite low; if randomly distributed in the genome, our markers would be separated by an average distance of about 12 Mb. This does not reflect a low number of markers screened as the ten primer combinations used here is comparable, if not greater than most previous studies (five primer pairs used by Williams & Oleksiak 2008 for example). Relatively few loci per primer pair were considered because we chose only high quality markers (clear presence/absence of the peak) and because AFLP profiles were very similar between *R. rattus* individuals; only about 35% of the observed bands were polymorphic. The low level of polymorphism of Malagasy *R. rattus* populations had already been noted based on mitochondrial sequences, and would result from the introduction of the black rat in Madagascar following human migration a few thousand years ago (Tollenaere *et al.* 2010). The number of markers necessary to detect selection in our study system is expected to be reduced by the presence of (i) a strong selective pressure (high fatality rate of plague in naive rodent populations) and (ii) population genetic bottlenecks on various spatiotemporal scales, first at the time of invasion of Madagascar by the species and second by regular population crashes following plague epidemics. Hitchhiking may thus involve large portions of the genome in Malagasy black rat populations.

The sampling design used in this study involved four independent pairs of populations experiencing different selection pressure. We identified loci as potentially under selection from higher differentiation (F_{ST}) than expected under neutrality between plague focus and plague-free zones. However, high F_{ST} values can result from other mechanisms such as hitchhiking in structured populations (Faure *et al.* 2008) or the founder effect (Excoffier & Ray 2008).

Neutral genetic differentiation between population pairs was low, especially in the population pair E2, maximizing the likelihood of detecting positively selected loci (higher F_{ST}, Beaumont 2005). On the other hand, the demographic histories of Malagasy *R. rattus*

Table 3 Results of selection detection for the 22 zone-specific outlier loci detected.

Outlier locus	DetSelD Populations pairs	Combined P-value	DFdist Population pairs	Combined P-value	Zone association q-value
A066	**W2**	0.009		0.404	0.59701
A121	E1, **E2**	0.001	**E1, E2**	0.000	0.00001
A284	E1, **W1**, **W2**	0.000	**W1**	0.006	0.03845
A319	E2	0.008	E2	0.232	0.89720
A449	W1, W2	0.005		0.211	0.00091
B161		0.505	E1, W2	0.003	0.00003
B163	**E1**	0.002	**E1**	0.009	0.00234
B408	**E1**	0.021	**E1**	0.002	0.00066
C411	E1, W1, **W2**	0.000	W2	0.046	0.00003
C483	**E1**, W2	0.001	**E1**	0.001	0.63067
D077	E1, W2	0.007	E1	0.108	0.00296
D180	**W2**	0.014	**W2**	0.001	0.00005
D450	E1	0.007	E1	0.105	0.00008
D455	E2	0.067	E1, E2	0.036	0.05437
D466	**E2**, W2	0.002	**E2**, W2	0.004	0.07913
J275	**E1**, W2	0.003	E1	0.056	0.00464
J393	**E2**	0.007		0.183	0.00464
Q199	**E1**, E2	0.006	**E1**	0.003	0.61325
U296		0.099	W1, **W2**	0.021	0.00096
U493	**W1**	0.012	**W1**, W2	0.012	0.00091
W099	E1, W2	0.004	W2	0.098	0.01128
Y222	E1, W2	0.005	W1, **W2**	0.005	0.00046

For each outlier locus, grey background indicates when the method used identified this locus. For repetition analyses, the significant population pairs are indicated. Population pairs are in bold when the locus revealed outlier at the 1% threshold. Combined P-values (Fisher's method) over the four independent comparisons are shown. Q-values (following Benjamini & Hochberg 1995) for the Fisher's exact test of association between the factor 'zone' (plague focus or plague-free zone) and the presence/absence of the AFLP peak are indicated.

Table 4 Analysis of molecular variance (AMOVA) grouping by zone (plague focus *vs.* plague-free zone) and by longitude location (east *vs.* west)

	'Plague focus' *vs.* 'Plague-free zone'				'East' *vs.* 'West'			
	df	Percentage of variation	P-value	statistics	df	Percentage of variation	P-value	statistics
a) Neutral dataset (227 loci)								
Among groups	1	0.22	0.20455	$\varphi_{CT} = 0.0022$	1	1.49	0.00000	$\varphi_{CT} = 0.0149$
Among populations within groups	6	4.80	0.00000	$\varphi_{SC} = 0.0481$	6	4.05	0.0000	$\varphi_{SC} = 0.0411$
Within populations	231	94.98	0.00000	$\varphi_{ST} = 0.0502$	231	94.46	0.00000	$\varphi_{ST} = 0.0554$
b) Outlier dataset (22 loci)								
Among groups	1	5.76	0.00000	$\varphi_{CT} = 0.0576$	1	–0.82	0.6004	$\varphi_{CT} = -0.008$
Among populations within groups	6	7.82	0.00000	$\varphi_{SC} = 0.0829$	6	11.9	0.00000	$\varphi_{SC} = 0.1180$
Within populations	231	86.43	0.00000	$\varphi_{ST} = 0.1357$	231	88.92	0.00000	$\varphi_{ST} = 0.1107$
c) Outlier sub-dataset (16 loci)								
Among groups	1	8.10	0.00000	$\varphi_{CT} = 0.0810$	1	–1.35	0.70812	$\varphi_{CT} = -0.0135$
Among populations within groups	6	7.37	0.00000	$\varphi_{SC} = 0.0802$	6	13.28	0.00000	$\varphi_{SC} = 0.1310$
Within populations	231	84.53	0.00000	$\varphi_{ST} = 0.1547$	231	88.07	0.00000	$\varphi_{ST} = 0.1193$

Analyses were performed with (a) the 227 neutral loci (b) the 22 zone-specific outliers (c) the 16 zone-specific outliers with significant association with the zone.

populations may not conform to demographic models used for simulations. *DetSelD* uses a pure divergence model, whereas migration may connect sampled

populations. *DFdist* includes migration by simulating a symmetrical island model, but dissimilar population size and asymmetric migration rate may occur as the

Table 5 Association between presence/absence of the peak corresponding to two AFLP markers (B408 and D466) and experimental plague challenge issue (survival or death following infection). The first dataset corresponds to a high-dose plague challenge in which 54 rats were injected with suspension containing 10^5 *Yersinia pestis*. The second dataset results from a low dose challenge with 49 animals injected with 100 *Y. pestis*. For each locus and each experiment, relative risks (RR) were calculated following Haldane and significance (*P*) was assessed by Fisher's exact test. We obtained combined *p*-values using the Fisher method on the two *P*-values of each experiment.

		B408 peak				D466 peak			
		Absence	Presence	RR	*P*	Absence	Presence	RR	*P*
High dose	Surviving	12(44.4%)	15(55.6%)	6.177	0.014	17(63.0%)	10(37.0%)	0.068	0.005
	Dying	3(11.1%)	24(88.9%)			26(96.3%)	1(3.7%)		
Low dose	Surviving	12(34.3%)	23(65.7%)	3.066	0.294	26(74.3%)	9(25.7%)	0.792	1
	Dying	2(14.3%)	12(85.7%)			11(78.6%)	3(21.4%)		
Combined *P*-values		*P* = 0.026				*P* = 0.032			

two zones differ in their catching rates (J.M. Duplantier, unpublished) and genetic diversities (in mitochondrial sequences, Tollenaere *et al.* 2010; in microsatellites, Brouat *et al.* unpublished). Nevertheless, the statistical methods used were shown to be relatively robust to departures from demographic assumptions (Beaumont & Nichols 1996; Vitalis *et al.* 2001). Moreover, the use of a design that compares population pairs partly avoids problems associated with heterogeneous demographic parameters (Tsakas & Krimbas 1976; Vitalis *et al.* 2001).

We chose to consider as outlier loci those detected in at least two population pairs as well as those having significant combined *p*-values for the four pairs in at least one simulation program (*DFdist* or *DetSelD*, each used with a threshold of 5%). This exhaustive method was chosen to detect outliers for two reasons: first, the reason that some loci are detected by only one simulation program is not fully understood (Beaumont 2005); second, dominant markers are less informative than codominant ones (Foll & Gaggiotti 2008). Moreover, we preferred error to occur through detection of false positives rather than missing important loci (Foll & Gaggiotti 2008) as the effects of selection can be confirmed from further analysis including sequencing and identification of outlier loci followed by molecular analysis of neighboring genes. Such subsequent analyses will be made easier by the phylogenetic proximity of the fully sequenced species *R. norvegicus* (Gibbs *et al.* 2004). Indeed, the two species diverged about 2.9 Myr ago (Robins *et al.* 2008), the proportion of variable sites between their mitochondrial genomes is 9.62% (estimated between *R. rattus*, Genbank Accession number: EU273707, and *R. norvegicus* Wistar strain, Genbank Accession number: NC_001665.2) and chromosome painting revealed complete homology and synteny between *R. rattus* and *R. norvegicus* in 16 of 18 chromosomes (Cavagna *et al.* 2002).

Using the detection method mentioned above, we found 22 candidate loci for zone-specific natural selection out of 249 total loci (8.8%). Sixteen of these loci had significant association with plague zone. Both datasets of outlier loci (22 or 16 loci) showed significant differentiation between plague-free zone and plague focus, but the subset of 16 loci better explained genetic variance (8.1% attributed to zone) than the 22 outlier loci (5.8%). On the other hand, neutral loci revealed significant east–west geographical differentiation, reflecting the relationship between gene flow levels and geographic distance under neutrality.

Under the hypothesis of a homogeneous selective pressure in the entire plague focus, we expected selected markers to be detected in all population pairs investigated; however, no locus fulfils such expectations. We chose as a criterion repetition of detection in at least two pairs and we also identified two loci (A284 and C411) that are highly differentiated in three out of four population pairs using *DetSelD*. Various hypotheses may explain the fact that outlier loci do not conform to all expected patterns, including restriction of evidence of selection to some of the four population pairs investigated or no association with plague zone of some outlier loci. For one thing, we cannot rule out the possibility of convergent evolution towards plague resistance in various populations of the Malagasy plague focus, potentially involving different genetic basis resulting in the same resistance phenotype. For example, adaptive melanism is controlled by different genes in two American populations of rock pocket mice (Hoekstra & Nachman 2003) and various epistatic networks mediate susceptibility to infection by a trypanosome in *Bombus terrestris* in Switzerland (Wilfert *et al.* 2007). An additional possibility is that adaptation through standing variation could lead to patterns of polymorphism different from the standard models of adaptation through new mutation and to the

association of functional polymorphisms with different alleles at closely linked loci (Barrett & Schluter 2008). It is also possible that, if variability in *Yersinia pestis* exists in Madagascar, coevolution may lead to variable resistance genetics in the host according to the locally present plague bacterium strain. *Yersinia pestis* strains from Madagascar all grouped in the cluster of strains that spread during the third pandemic, but some variability was described within the country (Guiyoule *et al.* 1997; Achtman *et al.* 2004). However, detailed analysis of *Y. pestis* genetic structure in Malagasy plague focus on which to base the investigation of such a hypothesis is lacking. Finally, plague-linked selective pressures may vary across space and time within the plague focus. A putative evolutionary cost of plague resistance may imply dramatic differences in the prevalence of resistance alleles in local populations that depend on the duration, time of occurrence and strength of the last epizootic event.

The outlier loci identified can be considered zone-specific. However, selective pressure cannot firmly be associated with plague, as other ecological factors differ between the two zones. Indeed, altitude and, as a consequence, mean temperature are confounded with plague occurrence (Jury 2003). On the other hand, annual precipitation (Jury 2003), vegetation (Lowry *et al.* 1997) and human culture (Battistini & Verin 1972; for example the type of houses, which may affect commensal rodents such as the black rat) differ between the eastern and western parts of the low-altitude zone. The nine outlier loci found in both eastern and western comparison pairs are thus good candidates for plague-specific selection because they are potentially affected by fewer confounding potential selective pressures.

We combined population genomics with a phenotype association study (Storz 2005; Stinchcombe & Hoekstra 2008) to discriminate between different possible selective pressures that differ between the central highlands and the low-altitude zone. Experimentally plague-challenged animals were typed for AFLP loci to test whether the plague challenge issue was correlated with outlier loci. Two outlier loci (B408 and D466) showed significant association with plague infection in one plague challenge experiment. The tendency remained the same in the second experiment, but the difference was not significant. This may be due to a lack of power of this second dataset (power of 63.8% for the locus B408 and 74.8% for D466 for an effect size equivalent to the one of the first dataset, data not shown) or to a lower discrimination of phenotypes by the low dose of bacteria threshold than by the high lose. These two loci were thus detected in two completely independent approaches and represent particularly relevant candidates for plague resistance markers.

The 20 outlier loci that were not associated with survival/death following plague challenge could, however, be related to plague resistance. First, the sample size of our plague challenge datasets (about 50 individuals) may be too low to detect modest genetic differences (Hirschhorn & Daly 2005). Second, the two doses of bacteria used may discern only poorly genetically differentiated groups. Third, we cannot rule out the possibility that a resistance genotype could be fixed in the central highlands. Indeed, the observation of phenotypic heterogeneity does not necessarily imply the presence of genetic polymorphism for plague resistance. If resistance genotype is fixed, the fact that some animals died after experimental challenge would in this case be related only to non-genetic factors such as immune and physiological status, as we used wild, not inbred, animals. Fourth, genetic differences associated with plague resistance/susceptibility may differ between plague focus and plague-free zone (outlier loci) and with respect to loci potentially detected through experimentally challenge of animals from the plague focus.

Conclusion and perspectives

This study identified 22 loci that represent candidates for zone-specific selection and that may be involved in the plague resistance phenotype of *R. rattus* populations in the central highlands of Madagascar. Two of these loci were also associated with plague challenge outcome in experimentally infected individuals. The major challenge now, and indeed a major issue for all studies of this type, is to move from the identification of an associated polymorphic AFLP locus to the identification of the underlying gene that is under selection. The strategy that is utilized is to isolate outlier AFLP bands then sequence these and use BLAST and equivalent analyses to identify their genomic location and neighbour genes. Such strategy is made possible by the current good information concerning the *R. norvegicus* genome (genome sequence and gene annotation), a species phylogenetically closed to *R. rattus*. Hence the potential for an unknown sequence to be identified is maximized. Preliminary sequencing analyses of the outlier loci identified in this study (10 markers out of 22, unpublished) are returning positive matches with the *R. norvegicus* genome, with some loci located closed to genes involved in the immune function, which appear intuitively biologically meaningful. Such genes will then be sequenced so that patterns of selection on DNA sequences can be investigated and functional polymorphism identified. As such, this study represents an important step in understanding the genetic basis of disease resistance and pathogen-mediated selection in natural populations of reservoirs.

Acknowledgements

References

Acevedo-Whitehouse K, Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology & Evolution*, **21**, 433–438.

Achtman M, Morelli G, Zhu PX *et al.* (2004) Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proceedings of the National Academy of Sciences, USA*, **101**, 17837–17842.

Ajmone-Marsan P, Valentini A, Cassandro M *et al.* (1997) AFLP (TM) markers for DNA fingerprinting in cattle. *Animal Genetics*, **28**, 418–426.

Altizer S, Harvell D, Friedle E (2003) Rapid evolutionary dynamics and disease threats to biodiversity. *Trends in Ecology and Evolution*, **18**, 589–596.

Andolfatto P (2001) Adaptive hitchhiking effects on genome variability. *Current Opinion in Genetics & Development*, **11**, 635–641.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, **23**, 38–44.

Battistini R, Verin P (1972) Man and the environment in Madagascar: past problems and problems of today. In:*Biogeography and Ecology in Madagasca* (eds Battistini R, Richard-Vindard G). pp. 311–337. Junk, The Hague.

Beaumont MA (2005) Adaptation and speciation: what F_{ST} can tell us? *Trends in Ecology and Evolution*, **20**, 435–440.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B*, **263**, 1619–1626.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, **57**, 289–300.

Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: Genome-wide sampling of insect populations. *Annual Review of Entomology*, **46**, 441–469.

Bonin A, Pompanon F, Taberlet P (2005) Use of amplified fragment length polymorphism (AFLP) markers in surveys of vertebrate diversity. In:*Molecular Evolution: Producing the Biochemical Data, Part B Methods in Enzymolog* (eds Zimmer EA, Roalson E). pp. 145–161. Academic Press, New York.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773–783.

Brygoo ER (1966) Epidémiologie de la peste à Madagascar. *Archives de l'Institut Pasteur de Madagascar*, **35**, 9–147.

Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**, 945–956.

Cavagna P, Stone G, Stanyon R (2002) Black rat (*Rattus rattus*) genomic variability characterized by chromosome painting. *Mammalian Genome*, **13**, 157–163.

Cavalli-Sforza LL (1966) Population structure and human evolution. *Proceedings of The Royal Society of London B*, **164**, 362–379.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155–174.

Daszak P, Cunningham AA, Hyatt AD (2000) Emerging infectious diseases of wildlife - threats to biodiversity and human health. *Science*, **287**, 443–449.

Dennis DT, Gage KL, Gratz N, Poland J, Tikhomirov E (1999) *Plague Manual: Epidemiology, Distribution, Surveillance and Control*. World Health Organisation, Geneva.

Duplantier J-M, Catalan J, Orth A, Grolleau B, Britton-Davidian J (2003) Systematics of the black rat in Madagascar: consequences for the transmission and distribution of plague. *Biological Journal of the Linnean Society*, **78**, 335–341.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature*, **452**, 169–175.

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, **23**, 347–351.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Faure MF, David P, Bonhomme F, Bierne N (2008) Genetic hitchhiking in a subdivided population of *Mytilus edulis*. *BMC Evolutionary Biology*, **8**, 164.

Fisher RA (1948) Combining independent tests of significance. *American Statistician*, **2**, 30.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Gage KL, Kosoy MY (2005) Natural History of plague: perspectives from more than a century of research. *Annual Review of Entomology*, **50**, 505–528.

Gibbs RA, Weinstock GM, Metzker ML *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **428**, 493–521.

Guiyoule A, Rasoamanana B, Buchrieser C *et al.* (1997) Recent emergence of new variants of *Yersinia pestis* in Madagascar. *Journal of Clinical Microbiology*, **35**, 2826–2833.

Haldane JBS (1956) The estimation and significance of the logarithm of a ratio of frequencies. *Annals of Human Genetics*, **20**, 309–311.

Hanson DA, Britten HB, Restani M, Washburn LR (2007) High prevalence of *Yersinia pestis* in black-tailed prairie dog

colonies during an apparent enzootic phase of sylvatic plague. *Conservation Genetics*, **8**, 789–795.

Herrera CM, Bazaga P (2008) Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet. *Molecular Ecology*, **17**, 5378–5390.

Hill AVS (2001) The genomics and genetics of human infectious disease susceptibility. *Annual Review of Genomics and Human Genetics*, **2**, 373–400.

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, **6**, 95–108.

Hoekstra HE, Nachman MW (2003) Different genes underlie adaptive melanism in different populations of rock pocket mice. *Molecular Ecology*, **12**, 1185–1194.

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.

Jury MR (2003) The climate of Madagascar. In:*The Natural History of Madagasca* (eds Goodman SM, Benstead JP). pp. 75–87. University of Chicago Press, Chicago.

Lazzaro BP (2008) Natural selection on the Drosophila antimicrobial immune system. *Current Opinion in Microbiology*, **11**, 284–289.

Lowry II PP, Schatz GE, Phillipson PB (1997) The classification of natural and anthropogenic vegetation in Madagascar. In:*Natural Change and Human Impact in Madagasca* (eds Goodman SM, Patterson BD). pp. 93–123. Smithsonian Institution Press, Washington and London.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science*, **12**, 106–117.

Meyer CL, Vitalis R, Saumitou-Laprade P, Castric V (2009) Genomic pattern of adaptive divergence in *Arabidopsis halleri*, a model species for tolerance to heavy metal. *Molecular Ecology*, **18**, 2050–2062.

Midamegbe A, Vitalis R, Malausa T, Streiff R (submitted). Scanning the European corn borer (*Ostrinia* spp.) for adaptive divergence between host affiliated sibling species.

Migliani R, Chanteau S, Rahalison L *et al.* (2006) Epidemiological trends for human plague in Madagascar during the second half of the 20th century: a survey of 20 900 notified cases. *Tropical Medicine and International Health*, **11**, 1228–1237.

Miller NJ, Ciosi M, Sappington TW *et al.* (2007) Genome scan of *Diabrotica virgifera virgifera* for genetic variation associated with crop rotation tolerance. *Journal of Applied Entomology*, **131**, 378–385.

Nielsen R, Bustamante C, Clark AG *et al.* (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *Plos Biology*, **3**, 976–985.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: 'Isolation by adaptation' and multiple roles for divergent selection. *Evolution*, **62**, 316–336.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, **96**, 7–21.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.

Rahalison L, Ranjalahy M, Duplantier J-M *et al.* (2003) Susceptibility to plague of the rodents in Antananarivo, Madagascar. In: *The Genus Yersini* (eds Skurnik M, Bengoechea JA, Granfors K). pp. 439–442. Kluwer Academic/Plenum Publishers, New York.

Robins JH, McLenachan PA, Phillips MJ *et al.* (2008) Dating of divergences within the Rattus genus phylogeny using whole mitochondrial genomes. *Molecular Phylogenetics and Evolution*, **49**, 460–466.

Schlotterer C (2003) Hitchhiking mapping—functional genomics from the population genetics perspective. *Trends in Genetics*, **19**, 32–38.

Sorci G, Moller AP, Boulinier T (1997) Genetics of host-parasite interactions. *Trends in Ecology and Evolution*, **12**, 196–200.

Stenseth NC, Atshabar BB, Begon M *et al.* (2008) Plague: past, present, and future. *Plos Medicine*, **5**, 9–13.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–170.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Storz JF, Hoekstra HE (2007) The study of adaptation and speciation in the genomic era. *Journal of Mammalogy*, **88**, 1–4.

Thomas RE, Barnes AM, Quan TJ *et al.* (1988) Susceptibility to Yersinia pestis in the northern grasshopper mouse (*Onychomys leucogaster*). *Journal of Wildlife Diseases*, **24**, 327–333.

Tollenaere C, Rahalison L, Ranjalahy M *et al.* (2008) CCR5 polymorphism and plague resistance in natural populations of the black rat in Madagascar. *Infection Genetics and Evolution*, **8**, 891–897.

Tollenaere C, Brouat C, Duplantier JM *et al.* (2010) Phylogeography of the introduced species Rattus rattus in the western Indian Ocean, with special emphasis on the colonization history of Madagascar. *Journal of Biogeography*, **37**, 398–410.

Tsakas S, Krimbas CB (1976) Testing the heterogeneity of F values: a suggestion and a correction. *Genetics*, **84**, 399–401.

Turner JK, McAllister MM, Xu JL, Tapping RI (2008) The resistance of BALB/cJ mice to *Yersinia pestis* maps to the major histocompatibility complex of chromosome 17. *Infection and Immunity*, **76**, 4092–4099.

Vekemans X (2002) *AFLP-surv version 1.0*. Distributed by the author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, **158**, 1811–1823.

Vitalis R, Dawson K, Boursot P, Belkhir K (2003) DetSel 1.0: A computer program to detect markers responding to selection. *Journal of Heredity*, **94**, 429–431.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Wilfert L, Schmid-Hempel P (2008) The genetic architecture of susceptibility to parasites. *Bmc Evolutionary Biology*, **8**, 187.

Wilfert L, Gadau J, Baer B, Schmid-Hempel P (2007) Natural variation in the genetic architecture of a host–parasite interaction in the bumblebee *Bombus terrestris*. *Molecular Ecology*, **16**, 1327–1339.

Williams LM, Oleksiak MF (2008) Signatures of selection in natural populations adapted to chronic pollution. *Bmc Evolutionary Biology*, **8**, 282.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, **8**, 907–913.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Results of outlier detection using *DetSelD* program for the four independent comparisons of one low altitude (plague-free zone) and one high altitude (plague focus) *R. rattus* populations. For each population pair, the differentiation of the second population to the ancestral population (F_2) is plotted against the differentiation of the first population to the ancestral population (F_1). The dark grey colored part represents the envelope containing 95% of simulations, whereas the light grey contains 99%. Each point corresponds to one locus. Neutral loci are plotted in black whereas those revealing higher differentiation are plotted with red (P-value < 0.01) or orange (P-value < 0.05) points.

Fig. S2 Results of outlier detection using *DFdist* program for the four independent comparisons of one low altitude (plague-free zone) and one high altitude (plague focus) *R. rattus* populations. For each population pair, the genetic differentiation between the two populations of the pair (F_{ST}) is plotted against expected heterozygosity (H_E). The solid line delimits the envelope containing 95% of simulations, whereas the dotted line defines 99%. Each point corresponds to one locus. Neutral loci are plotted in black whereas those revealing higher differentiation are plotted with red (P-value < 0.01) or orange (P-value < 0.05) points.