Differential immune gene expression associated with contemporary range expansion of two invasive rodents in Senegal

Nathalie Charbonnel[1*], Maxime Galan[1], Caroline Tatard[1], Anne Loiseau[1], Christophe Diagne[2], Ambroise Dalecky[4], Hughes Parinello[5], Stephanie Rialle[5], Dany Severac[5] and Carine Brouat[2]

[1] CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Univ Montpellier, Montpellier, France

[2] CBGP, IRD, CIRAD, INRA, Montpellier SupAgro, Univ Montpellier, Montpellier, France

[3] Dept de Biologie Animale, Faculté des Sciences et Techniques, Université Cheikh Anta Diop (UCAD), Fann, Dakar, Senegal

[4] IRD, Aix Marseille Univ, LPED, Marseille, France

[5] MGX-Montpellier GenomiX, c/o Institut de Génomique Fonctionnelle, Montpellier, France

Abstract

Background: Biological invasions are major anthropogenic changes associated with threats to biodiversity and health. What determines the successful establishment of introduced populations still remains unsolved. Here we explore the appealing assertion that invasion success relies on immune phenotypic traits that would be advantageous in recently invaded sites.

Results: We compared gene expression profiles between anciently and recently established populations of two major invading species, the house mouse *Mus musculus domesticus* and the black rat *Rattus rattus*, in Senegal. Transcriptome analyses revealed respectively 364 and 83 differentially expressed genes along the mouse and rat invasion routes. Among them,

20.0% and 10.6% were annotated with functions related to immunity. All immune-related genes detected along the mouse invasion route were over-expressed in recently invaded sites. Genes of the complement activation pathway were over-represented. Results were less straightforward when considering the black rat as no particular immunological process was over-represented.

Conclusions: We revealed changes in transcriptome profiles along invasion routes. Patterns differed between invaders. They could potentially be driven by increased infection risks in recently invaded sites for the house mouse and stochastic events associated with colonization history for the black rat. These results provide a first step in identifying the immune ecoevolutionary processes potentially involved in invasion success.

Keywords: Differential gene expression, evolutionary genomics, immunity, invasion, rodents, transcriptomics

Background

Biological invasions are considered as a major component of global change growing threats to biodiversity [1], ecosystem functioning [2] and human health [3]. They may also lead to huge negative economic impacts [4]. There is hence an urgent need for understanding what determines invasion success (*i.e.* the successful introduction, establishment and spread of species outside their native range) to design effective prevention, control and management strategies.

From an eco-evolutionary perspective, invasion success may rely on pre-adaptation within the original range [5,6] or on the rapid evolution of phenotypic traits that would be advantageous in newly colonized areas [5,7]. Some supports for this latter process come from the identification of phenotypic variation along invasion gradients, what has been evidenced in various animal species [7]. In this context, a corpus of appealing hypotheses has focused on traits that are related with the ability to cope with novel environments [8]. In particular, important changes of parasite pressures [8,9] can occur on introduced species along invasion gradients (e.g. enemy release, spill-over of native parasites [10-12]). Thus, invasion success could rely on the efficiency of immune system regulation, either through the reallocation of the energetic resources dedicated to defences towards other life history traits favouring range expansion (including dispersal or reproduction, evolution of increased competitive ability (EICA) hypothesis [13]), or by the modulation of defence strategies to limit the costs of immunity (EICA-refined hypothesis [14]). A recent analysis reviewing empirical studies has shown that in most cases, variations in immune responses were observed during invasions although no general pattern could be emphasized to corroborate the EICA hypotheses [15]. One of the problems pointed out by the authors was the lack of studies providing a large array of immune responses to properly assess immunocompetence and estimate immune trade-offs. Unfortunately, immune phenotyping often relies on a small number of immune effectors, because of the low quantity

of materials available (blood, tissue), or the difficulty to get immune kits for non-model species.

The recent advent of 'omics' technologies provides opportunities to explore a wide set of genes and pathways involved in the response to novel conditions experienced at invasion fronts. Genome scans have been successfully performed to study some invasion cases (e.g. [16-18]). They have enabled to detect signatures of contemporary adaptation associated with invasion, some of them being related to immunity [19]. Nevertheless, this genomic approach based on the detection of single nucleotide polymorphism outliers does not enable to identify the molecular processes driving the rapid adaptation that might occur within few generations. It also prevents from catching the importance of phenotypic plasticity in invasion success. Transcriptomics, the analysis of gene expression at a genome wide scale, is a complementary approach that might allow filling these gaps and deciphering the molecular mechanisms that underlie phenotypic changes [20,21]. Indeed, gene expression is an essential mechanism for rapid acclimatization or adaptation to novel environments [22-26]. Transcriptomics has hence been applied to identify genes and gene regulatory pathways that contribute to phenotypic variation at invasion fronts. Evidence for transcriptomic divergence along invasion gradients have already been reported [18,27], sometimes emphasizing the importance of immune regulation in invasion success [28]. However there is yet no study combining analyses of immune gene expression and potential variations in parasite pressures along invasion routes. These two elements are essential to test the EICA hypotheses and further understand the interactions between immune system regulation and invasion success [15].

Here, we examined transcriptional profiles along invasion routes of the house mouse *Mus musculus domesticus* and the black rat *Rattus rattus* currently invading Senegal (West Africa). These two rodent species have been colonizing Senegal eastwards from western coastal colonial cities since the beginning of the 20th century [29,30]. Our previous study based on few

immune effectors highlighted higher inflammatory and/or antibody-mediated responses in recently invaded sites compared with anciently invaded ones for both invaders [31]. These immune variations could be considered as responses to novel parasite pressures encountered in recently invaded areas. Surveys of bacterial communities in Senegal seemed to corroborate this prediction [32] as both quantitative and qualitative changes in bacterial composition were observed along the house mouse and black rat routes of invasion. However this pattern could not be generalised to any 'parasite' pressure as a decrease of the overall prevalence in gastrointestinal helminths was also observed, suggesting enemy release for these macroparasites [33]. We therefore developed an RNAseq approach to assess without any *a priori* differential gene expression patterns between populations from anciently and recently invaded sites. We tested the null prediction that immune gene expression patterns would not differ along invasion routes. We investigated two alternative hypotheses. On one hand, we expected an overall higher immune gene expression of rodent populations in recently invaded sites, as a response to novel parasite pressures encountered [31]. On the other hand, we expected trade-offs between immune pathways in recently invaded sites under the EICA-refined hypothesis. Enemy release of macroparasites and changes in the composition of bacterial communities in recently invaded sites could mediate such trade-offs. More specifically, we investigated the following questions: (i) Are there differences in immune gene expression between anciently invaded and recently invaded populations of mice and rats? (ii) Are there functional categories of immune genes that are over-represented among the differentially expressed genes? iii) Do these results support the EICA or EICA-refined hypotheses?

Results

Qualitative description of expression patterns

The 32 transcriptome libraries sequenced (see Table 1 for details) produced an average of 51,6 million read pairs. The numbers of retained reads post-filtering ranged from 37.2 M to 76.7 M, corresponding to 92.82% of the reads. When considering the house mouse data (3 mismatches/read pair allowed), an average of 85.68 % of the quality-filtered reads mapped to the *M. musculus domesticus* reference genome and 1.58% of these latter mapped to multiple regions. When considering the black rat *R. rattus* data and the possibility for 9 mismatches/read pair with *R. norvegicus* reference genome, 81.66% of the quality-filtered reads were mapped to the reference genome of the related *R. norvegicus*. Among them, 0.98% mapped to multiple regions of the reference genome.

The principal component analysis (PCA) performed on the house mouse standardized read counts separated sampling sites along the second principal component, which accounted for 4.58% of the variance observed (Suppl. Fig. S1a). We observed a slight opposition between sites corresponding to anciently and recently invaded areas, except for Aere Lao that did not group with recently invaded sites. No separation of sites according to invasion-related categories was visible on the PCA performed on the black rat data. On the second principal component that represented 2.98 % of the variance explained, two sites from anciently invaded areas were opposed to the six other ones (Suppl. Fig. S1b).

Differential expression (DE) analyses

For both datasets, a high dispersion of the total number of reads could be observed between biological replicates (Suppl. Fig. S2a,b). Therefore we applied differential expression (DE) analyses using two strategies. The first one named '4vs4 approach' considered a single randomly chosen biological replicate for each site and analysed the 256 comparisons of the relative log expression (RLE)-transformed normalized read counts between anciently and recently invaded sites. This approach is very conservative. The second one named '8vs8

approach' considered the eight possibilities of 'site x replicate' and employed a generalized linear model (GLM) to take into account the fact that each site was represented by two biological replicates in this analysis. This approach is less conservative.

Along the mouse invasion route, over the 17,853 filtered genes (more than 10 occurrences) tested using edgeR and the '4vs4 approach', 18 genes were always differentially expressed between anciently and recently invaded sites. Over the 16,630 filtered genes (more than 10 occurrences) tested using the '8vs8 approach', we detected 593 differentially expressed genes among which 364 were found in at least 85% of the 256 previous '4vs4 approach' comparisons. Also note that only five genes were detected by both '4vs4' and '8vs8' approaches (Hal, Rnf183, Serpina6, Wif1, 9030619P08Rik; see details in Suppl. Table S1a).

Along the black rat invasion route, 54 genes were always differentially expressed between anciently and recently invaded sites among the 13,190 ones (more than 10 occurrences) analysed when using the '4vs4 approach'. We found 268 DE genes among the 12,747 ones (more than 10 occurrences) analysed with the '8vs8 approach' among which 83 were detected in at least 85% of the '4vs4' comparisons. Forty-two differentially expressed genes were common to the '4vs4' and '8vs8' approaches. Results are detailed in Suppl. Table S1b.

We have focused further analyses on the 364 and 83 genes that were identified as DE in the '8vs8' approach and detected in at least 85% of the '4vs4 'comparisons, for the house mouse and black rat, respectively, which we qualified as "robust" genes. The magnitude of expression differences were significantly different between the mouse and rat, with higher 'log fold changes' observed for mouse genes (Fig. 1, Wilcoxon test, $p = 0.01$). The log fold change (LogFC) per gene ranged between -9.04 and 2.34 for the mouse and between -5.79 and 4.49 for the rat robust genes. The average coverage varied from -3.16 to 11.94 LogCPM (counts per million) for the mouse and -2.41 to 9.56 for the rat robust genes.

Functional analyses of DE genes

Considering the house mouse dataset, STRING tool provided at least one gene ontology (GO) annotation for 345 of the 364 DE genes (Suppl. Table S1). Among the 2,098 GO annotations found, 357 GO terms corresponding to biological processes were significantly enriched (FDR < 0.05) when compared to the house mouse reference genome. In particular, 10 immune-related functions were significantly enriched, including inflammatory (GO:0006954), humoral (GO:0002455, GO:0006959), and innate (GO:0045087, GO:0034097) responses or activation/regulation of immunity (GO:0006952, GO:0002253, GO:0050776). We also detected enrichment for two biological processes related to response to wounding (GO:0009611, GO:1903035). These enriched GO terms related to immunity corresponded to 73 DE genes, i.e. 20% of all DE genes (see Suppl. Table S1). When taking into account redundancies using REVIGO (SimRel semantic similarity measure, similarity cutoff = 0.7), we found 20 main clusters for the category 'Biological Process'. Inflammation and acute-phase response were part of the second most represented one (Fig. 2). Using KEGG algorithm, we detected 47 enriched pathways. We found that 29 immune related genes belonged to these over-represented KEGG biological pathways. The most significant one being the immune 'Complement and coagulation cascades' (FDR = 1.56 e-30).

Interestingly, the distribution of average coverage (LogCPM) for these immune-related genes corresponding to enriched GO terms or pathways was shifted towards higher values compared to the 364 robust genes (Wilcoxon test, $p = 1.52 \times 10^{-3}$; Fig. 3, Suppl. Fig. S3). The heatmaps built on the normalized read counts for the immune related genes belonging to the biological processes (as well as pathways) that were found to be over-represented revealed a down-regulation of these genes for all anciently invaded sites and an up-regulation for all recently invaded sites except Aere Lao (Fig. 3, Suppl. Fig. S4). In this latter population, genes exhibited similar expression patterns than anciently invaded sites.

Finally, using STRING, we found that the protein-protein interaction (PPI) network based on the 364 DE genes was significantly enriched (p < 1.0e-16, 731 edges found, 54 expected), meaning that proteins have more interactions among themselves than what would be expected for a random set of proteins of similar size, drawn from the mouse genome. Two main clusters were detected, one including Cytochrome P450 and UDP glucuronosyltransferase proteins and the other being mainly composed of immune related proteins, including fibrinogen, serin peptidase inhibitor (serpin), apolipoprotein and complement proteins. The PPI interaction network built upon the 73 immune related genes belonging to the biological processes previously found to be over-represented was also statistically significant (p < 1.0e-16, 561 edges found, 13 expected). The main cluster included the fibrinogen and serine peptidase inhibitor (serpin) proteins. Other highly connected proteins included apolipoprotein A and haptoglobin on one hand, and complement proteins on the other hand (Fig. 4).

Considering the *R. rattus* dataset, GO annotations were available for 75 of the 83 DE genes in black rat (Suppl. Table S2). Eight of these annotated DE genes were related to immune biological processes (i.e. 10.66 % of all DE genes; Suppl. Table S2). We did not find any biological process or pathway significantly enriched, even when considering redundancies using REVIGO. The protein-protein interaction network built from STRING was not significantly enriched (p = 0.09).

Discussion

In this study, we presented a comparative transcriptomics approach based on spleen RNA sequencing for two invasive rodent species studied along their invasion routes in Senegal. We aimed at identifying adaptive divergence between recently and anciently established rodent

populations that may at least partly explain invasion success. More specifically, we investigated differential immune related gene expression that could provide support in favour of the EICA hypotheses. We observed differentially expressed genes along the mouse and rat invasion routes. Among them, 20.0% for the mouse and 10.6% for the rat were annotated with functions related to immunity. None of the results found with regard to these genes supported one of the EICA hypotheses. Contrary to what was expected, along the mouse invasion route, all immune-related genes detected were over-expressed in recently invaded sites, and among them, inflammation and complement pathways were over-represented. Considering the black rat, results were less straightforward - changes of expression levels were of lower amplitude than for the mouse genes - and no marked pattern could be drawn. Genes found to be differentially expressed were either up- or down-regulated in recently-invaded sites, but no particular biological process was significantly over-represented. Although bioinformatics limitations could underlie these differences between the house mouse and black rat (transcriptome annotation was less efficient for *R. rattus* as we used *R. norvegicus* transcriptome as a reference), our results suggested that variations of immune phenotypes were a less important strategy for *R. rattus* invasion success. Overall, it is likely that different processes, including colonization history and/or alternative mechanisms by which species adapt to novel environments, may have mediated the invasion success of these two rodent species. In the future, it could be interesting to take into account the potential variability of these patterns with regard to the targeted tissue or organ. We here focused on rodent spleen. Results would perhaps have been different if we had targeted other lymphoid organs, lymphatic tissues or non-immune targets.

Invasion success is supposed to rely on the evolution of phenotypic traits that would be advantageous at invasion fronts, where species are likely to be exposed to novel

environmental conditions [7,9,34]. Here, using comparative transcriptomics in wild populations, we have evidenced phenotypic variation along the mouse and rat invasion routes in Senegal. Indeed, we detected variations of gene expression levels, a phenotypic characterisation that results from a combination of genotype, environment and genotype–environment interactions [35]. This phenotypic variation may result from stochastic changes due to population history, including founder events and genetic drift [7]. Previous population genetic studies suggested serial founder events along invasion routes for both the house mouse [36] and the black rat [30] in Senegal. For the black rat, no particular biological process was significantly over-represented using this comparative transcriptomics approach. This absence of clear pattern suggested that stochastic events may have shaped the phenotypic variations observed along *R. rattus* invasion route. However, they are less plausible to explain the over-expression of immune genes detected in *M. musculus domesticus* recently invaded sites.

Variations of gene expression levels may also result from plastic responses to novel selection pressures (e.g. [37]) or from deterministic shifts involving natural selection and adaptation to local conditions [38]. Although mouse and rat populations experienced reduced genetic diversity due to founder events, they may have developed adaptive responses to novel selection pressures through high levels of plasticity [39,40]. In the house mouse, the functional categories of genes that were over-represented among the differentially expressed ones gave us insights into the phenotypic traits that may be concerned by deterministic shifts during invasion. Transcriptome sequencing revealed strong evidence for significantly biased variations in immune gene expression towards an up-regulation of complement and pro-inflammatory cascades in recently invaded sites. These results corroborated our previous functional studies [31]: immune challenges revealed that antibody-mediated (natural antibodies and complement) and inflammatory (haptoglobin) responses were increased in mice sampled in recently

invaded sites compared to those sampled in anciently invaded ones. This study was therefore in line with the hypothesis that parasite-mediated selection could contribute to phenotypic differentiation along invasion routes, potentially in interaction with other environmental factors. However, it still did not support the EICA or EICA refined hypotheses [14]. Despite the enemy release patterns detected for mouse helminths in Senegal (decrease of specific richness and of overall prevalence in recently invaded sites, see [33]), we did not find any evidence of immunity or particular immune pathways being dampened at the expense of other life history traits or of less energetically costly immune defences. The up-regulation of all immune genes found to be differentially expressed in sites recently invaded by the house mouse strongly supported the assumption of an increased overall infection risk in recently invaded sites. The changes in bacterial communities observed between anciently and recently invaded sites introduction of exotic bacterial genera and/or changes in the prevalence of endemic ones, see [32] could be in line with this potential increased risk. Similar rapid changes in immunity have been observed in other invading organisms, but these studies revealed a wide array of patterns that did not enable to conclude on the potential genericity / specificity of the EICA hypotheses [15]. We advocate for further studies using wide ranges of immune effectors, including 'omics methods' [41], to improve our understanding of the relationships between biological invasion and immunity.

Interestingly, the site of Aere Lao did not provide the expected signature along the mouse invasion route, as it provided similar results than anciently invaded sites. The difference of Aere Lao among recently-invaded sites was already noticeable in some of the effectors (complement; haptoglobin) previously measured by immune challenges [31]. This pattern may not be related to parasitological data, because endemic parasites were found at high levels of prevalence in this site [32,33]. On the contrary, it may reflect a more ancient invasion history of the house mouse in this locality than in the neighbouring ones, for which we have historical

data [29]. Indeed, Aere Lao is a bigger city than Dodel, Croisement Boubé or Lougue (P. Handschumacher, pers. com.) and a market place that may have been colonised earlier by the house mouse.

Concerning *R. rattus*, no particular pathway (including immune related ones) was significantly over-represented. In addition, only eight immune related genes were found to be differentially expressed. They encode for proteins that interact with bacteria (*Rnase6*, *Ptprz1*), viruses (*Ltc4s*, *Hspa1b*) or parasites (*Gfi1b*). Some of them were down-regulated in recently invaded sites (e.g. *Rnase6)* while others were strongly up-regulated (e.g. *Ptprz1*). So, no specific pattern emerged with regard to potential variations in infection risks. Interestingly, we also previously found from a functional immune study that differences in immune responses between the rat anciently and recently invaded sites were less marked than for the mouse [31]. In addition, we could hardly identify other biological processes that could reflect life history traits adaptation with regard to invasion. Those processes that were found to be associated with differentially expressed genes covered a wide range of functions including clustering of voltage sodium channels, negative regulation of cell-matrix adhesion and protein dephosphorylation and response to cyclic Adenosine Monophosphate (cAMP). This latter function could be interesting with regard to the rat invasion success as cAMP is central to the regulation of insulin and glucagon secretion [42]. Variations in cAMP gene expression could therefore mediate differences in response to stressful situations, including starvation or fight-or-flight response [43]. It would be interesting to analyse whether such differences could result in different behavioral phenotypes between rats trapped in anciently invaded sites (expected to exhibit little performance and high stress response in novel environments) and recently invaded ones (expected to exxhibit high response capacity in novel environments) [44]. Therefore it seems even more important to further perform transcriptomics analyses on other

organs and tissues to identify the phenotypic changes and ecoevolutionary processes linked with the black rat invasion success. Brain, where key genes underlying behavioral invasion syndrome are expected to be expressed, could be a relevant organ to target in the future [45].

In conclusion, our work revealed changes in transcriptomic profiles along invasion routes for both the house mouse and the black rat. The patterns observed could potentially be driven by increased infection risks in recently invaded sites for the house mouse and by stochastic changes for the black rat. Whether differences in expression were driven by phenotypic plasticity or directional selection during or after invasion, or reflected the colonization history of rodents, has yet to be confirmed, and should deserve genomic studies and experimental work [18,21]. Moreover, it would also be important to validate whether these changes in phenotypic traits influenced ecological dynamics (e.g. [46]) and in turn, invasion success.

Methods

Ethical statements

Sampling campaigns within private properties were realized with the prior explicit agreement from relevant familial, local and institutional authorities. None of the rodent species investigated here has protected status (see list of the International Union for Conservation of Nature). All animal-related procedures were performed according to official ethical guidelines provided by the American Society of Mammalogists [47]. All protocols presented here were realized with prior explicit agreement from relevant institutional committee (CBGP: D 34-169-1). They were carried out in accordance with requirements of Senegalese and French legislations.

Sampling sites, sample collection and RNA extraction

We used some of the spleen samples of *M. m. domesticus* and *R. rattus* that were previously analysed for the bacterial metabarcoding study previously described [32]. Briefly, for both invasive species, rodent trapping was performed in four anciently invaded (more than 100 years) and four recently invaded (less than 30 years) sites defined according to historical and longitudinal surveys [29]. Autopsies were realized at approximately the same period of the day for all individuals (11 am – 14 pm). Spleens were collected and immediately placed in RNAlater, stored at 4 °C overnight then at −20 °C until further analyses.

Twenty adult rodents per site were considered excepted for five sites where only 14-18 samples were available (see Table 1). Total RNA was extracted from approximately 5mg of spleen for each sample using AllPrep 96 DNA/RNA Kit (Qiagen). The quality and quantity of the purified RNA was assessed by gel electrophoresis and NanoDrop spectrophotometer (Thermo Scientific) before pooling, and Bioanalyzer 2100 (Agilent) after pooling (see below). For each rodent species, sixteen equimolar pools (2 per sites) that combined 6 to 10 individual samples with equilibrated sex ratio were made.

cDNA library preparation and RNA sequencing

cDNA library construction, sequencing and de novo assembly of the filtered reads were performed at MGX platform. RNA-Seq libraries were constructed with the Truseq stranded mRNA sample preparation (Low throughput protocol) kit from Illumina. One microgram of total RNA was used for the construction of the libraries. The first step in the workflow involved purifying the poly-A containing mRNA molecules using poly-T oligo attached magnetic beads (Poly-A based mRNA enrichment). Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using Actinomycine D, random hexamer primers and SuperScript II reverse transcriptase for *M. m. domesticus* samples or SuperScript

IV reverse transcriptase for *R. rattus* samples. The second strand cDNA was synthesized by replacing dTTP with dUTP. These cDNA fragments then have the addition of a single 'A' base and subsequent ligation of the adapter. The products were then purified and enriched with 15 cycles of PCR. The final cDNA libraries were validated and quantified with a KAPA qPCR kit.

For *M. m. domesticus* samples, four DNA libraries per sequencing line were pooled in equal proportions, denatured with NaOH and diluted to 7 pM before clustering. Cluster formation on a flowcell V3, primer hybridisation and single-end read 50 cycles sequencing were performed on cBot and HiSeq2000 (Illumina, San Diego, CA) respectively.

For *R. rattus* samples, four DNA libraries per sequencing line were pooled in equal proportions, denatured with NaOH and diluted to 8 pM before clustering. Cluster formation on a flowcell V4, primer hybridisation and single-end read 100 cycles sequencing were performed on cBot and HiSeq2500 (Illumina, San Diego, CA) respectively. We performed single-end read 100 cycles for *R. rattus* as up to known the only reference genome available for rats is from another species *Rattus norvegicus*.

Transcriptome analysis

Transcript assembly

Image analyses and base calling were performed using the Illumina HiSeq Control Software and Real-Time Analysis component. Demultiplexing and production of Fastq files were performed using Illumina's conversion software (CASAVA 1.8.2 for the house mouse data, BCL2FASTQ 2.17 for the black rat data). The quality of the raw data was assessed using FASTQC from the Babraham Institute and the Illumina software SAV (Sequencing Analysis Viewer).

A splice junction mapper, TOPHAT [48] (v2.0.9 for the mouse data, v2.0.13 for the black rat data), using BOWTIE [49] (v2.1.0 for the mouse data, v2.2.3 for the black rat data), was used to align the reads to the *M. musculus* genome (UCSC mm10) or the *Rattus norvegicus* genome (UCSC rn4) with a set of gene model annotations (genes.gtf downloaded from UCSC on March 6, 2013 for the mouse genome, on July 17, 2015 for the brown rat genome). Final read alignments having more than 3 mismatches (house mouse) or 9 mismatches (black rat) were discarded. Again we used a different threshold for mouse and rat data as the reference genome available for this latter is from *Rattus norvegicus*. Read alignment rates were above 83,3 % for all *M. m. domesticus* libraries and 74.0 % for all *R. rattus* libraries.

SAMTOOLS (v0.1.18 for the mouse data, v1.2 for the black rat data) was used to sort the alignment files. Then, gene counting was performed with HTSEQ count (v0.5.4p5 for the mouse data, v0.6.1p1 for the black rat data) using the union mode [50]. The data is from a strand-specific assay, the read has to be mapped to the opposite strand of the gene.

Analysis of differential gene expression between populations

Differentially expressed (DE) genes were identified using the BIOCONDUCTOR [51] package edgeR v3.4.0, under R 3.0.2 for the mouse data, v3.8.6 under R 3.2.3 for the black rat data, [52]. It enabled to take into account the potentially high variability observed between biological replicates. Data were normalized using the relative log expression Rle normalization factors [53]. Two analysis approaches have been used to identify the DE genes. First, one out of the two locality replicates has been randomly selected, in order to have 4 replicates in each condition ('recently invaded' versus 'anciently invaded' sites, '4vs4 approach'). In total, 256 combinations of samples were possible and the comparisons performed. Second, the eight samples from each condition were kept, and a 'locality' factor was added to the design in order to consider that replicates were paired ('8vs8 approach'). Before statistical analysis,

genes with less than 20 reads (first approach) or 40 reads (second approach), cumulating all the 8 or 16 analysed samples per species, were filtered and thus removed. It enabled to limit the number of statistical tests and therefore to lower the impact of multiple tests corrections. In these two approaches, the *p*-value threshold was set to 5%, after application of the Benjamini-Hochberg method for multiple testing correction.

The differentially expressed genes identified with the '8vs8 approach' and found in more than 85% of the 256 combinations of the '4vs4 approach' were declared as DE. The threshold of 85% was set according to the visualisation of a barplot representing the number of '4vs4' comparisons in which a gene (highlighted in the '8vs8' approach) is found to be differentially expressed.

We performed a principal component analysis (PCA) on normalized read counts of expressed genes (excluding weakly expressed ones, see above) using R v3.0.1 [54] and the FACTOMINER libray (PCA function). It enabled to visualize differences in gene expression between sites and biological replicates.

Functional and enrichment analyses

M. m. domesticus and *R. rattus* genes were functionally annotated with gene ontology (GO) terms using Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) v10.5 tool [55,56]. We performed gene enrichment analyses using Fisher's exact tests on the house mouse and black rat declared DE gene sets using house mouse *M. musculus* and brown rat *R. norvegicus* GO as the background and the software available in STRING. *p*-values were corrected for multiple testing using the Benjamini and Hochberg false discovery rate (FDR). GO terms with a FDR *p*-value < 0.05 were considered as statistically significantly different. The redundancies of significantly enriched GO terms were reduced using REVIGO [57] with a similarity cutoff of 0.7. Similarly we performed enrichment analyses on biological processes and biological pathways supplied by KEGG (http://www.genome.jp/kegg/pathway.html),

implemented in STRING.

Finally we performed a global protein network analysis based on DE genes using STRING database that included interaction databases (including direct (physical) as well as indirect (functional) associations), genetic interactions and shared pathway interactions. It enabled to show how key components of different pathways interact. Two PPI networks were constructed by mapping all DE genes and immune related DE genes, respectively, to STRING database with confidence score >0.7 (high level). PPI networks were visualized and analyzed in CYTOSCAPE software through the web interface of STRING. Enrichment in these protein-protein interactions were tested using Fisher's exact tests and correction for multiple testing (FDR).

Acknowledgements

We thank Khalilou Bâ, Mamoudou Diallo, Mamadou Kane, Laurent Granjon, Youssoupha Niang, Aliou Sow and Philippe Gauthier for their help in collecting samples, and J.-M. Duplantier for invaluable advice regarding field work and rodent invasion history. We are indebted to all the Senegalese people who allowed us to trap rodents in their homes. This work was supported by the ANR ENEMI project (ANR-11-JSV7-0006), and by the French Embassy in Senegal, which provided the funding for PhD scholarships.

Author Contributions

CB, MG and NC conceived the study. AD, CB, CAD, NC, participated in the population sampling (other people involved have already coauthored other published papers). MG, CT and AL performed the RNA extraction. HP and DS constructed the RNA libraries and performed the HiSeq sequencing. HP and SR performed bioinformatics and statistical

analyses for DE gene detection. NC conducted functional and enrichment analyses. NC wrote the manuscript, with contribution gathered from all co-authors.

Additional Information (including a Competing Financial Interests Statement)

We have no conflict of interest to declare.

Data availability

All sequence data, including raw reads and assemblies have been deposited on GEO. Gene ontologies for DE genes are provided in Supplementary Tables 1 and 2. Differential expression data are available from the corresponding author on request.

Supplementary information

Figure S1. Principal Components Analysis (PCA) of RLE-normalized read counts among samples partitioned by sites for a) *M. musculus domesticus* (16,630 genes) and b) *R. rattus* (12,747 genes) invasion routes.

Figure S2. Pairwise comparisons of the RLE-normalized total number of reads between libraries for a) the house mouse and b) the black rat. For the house mouse, anciently invaded sites are : Da=Dagathie ; Mb=Mbakhana, Th=Thilene, Nd=Ndombo; recently invaded sites are Cr=Croisement Boube, Do=Dodel, Ae=Aere Lao, Lo=Lougue. For the black rat, anciently invaded sites are : Diak=Diakene Wolof, Diat=Diattacounda, Mar=Marsassoum, Tob=Tobor ; recently invaded sites are : Bad=Badi Nieriko, Bou=Boutougoufara, Ked=Kedougou, Sou=Soutouta.

Figure S3. Boxplot representing the level of differential expression (in log fold change (log FC)) for immune related and non immune related robust genes identified along the mouse invasion route.

Figure S4. Heatmap of the DE genes between the anciently and recently invaded sites of *M. musculus domesticus*. The normalized read counts for the expressed genes are shown. Heatmap was built in R using heatmap.2 for a) all DE genes and b) 73 immune-related genes belonging to over-represented biological pathways. The genes (rows) and samples (columns) were clustered using dendrograms built with Ward distance and hierarchical clustering.

Table S1. a) Details of the 18 genes found to be differentially expressed between the anciently and recently invaded sites of *M. musculus domesticus* invasion route using the 4x4 approach. Genes indicated in bold were found to be differentially expressed with both 8x8 and 4x4 approaches. b) Details of the 364 genes found to be differentially expressed between the anciently and recently invaded sites of *M. musculus domesticus* invasion route using the 8x8 approach and found in 85% of the comparisons made using the 4x4 approach. The 73 genes indicated in red are related with immunity and have over-represented GO annotations. The 29 genes underlined and in italics correspond to the immune-related genes with over represented Kegg biological pathways. GO ID corresponds to Gene ontology annotations.

Table S2. a) Details of the 54 genes found to be differentially expressed between the anciently and recently invaded sites of *R. rattus* invasion route using the 4x4 approach. Genes indicated in bold were found to be differentially expressed with both 8x8 and 4x4 approaches. b) Details of the 83 genes found to be differentially expressed between the anciently and recently invaded sites of *R. rattus* invasion route using the 8x8 approach and found in 85% of the comparisons made using the 4x4 approach. Genes indicated in bold were found to be

differentially expressed with both 8x8 and 4x4 approaches. Genes indicated in red are related with immunity. GO ID corresponds to Gene ontology annotations.

Figure Legends

Figure 1. Significantly differentially expressed genes (orange dots) and immune related genes (red dots) between anciently invaded and recently invaded areas along a) the house mouse and b) the black rat invasion roads. The 73 immune related genes belonging to biological processes found to be significantly enriched are represented for the house mouse. Vertical lines indicate 1 log fold change (Log FC). The x-axis indicates genes that are down- (negative values) and up-regulated (positive values) in anciently compared to recently invaded areas. The y-axis represents the average level of gene expression using read counts per million (Log CPM).

Figure 2. TreeMap view of REVIGO Biological process analyses for the house mouse differentially expressed (DE) genes. Each rectangle represents a single cluster, that are grouped into 'superclusters' of related terms, represented with different colors. The size of the rectangles reflects the frequency of the Gene Ontology (GO) term in this set of DE genes.

Figure 3. Heatmap of the differentially expressed (DE) genes between the anciently and recently invaded sites of *M. musculus domesticus*. The normalized read counts for the expressed genes are shown. For clarity, the heatmap was built in R using heatmap.2 for 29 immune related genes belonging to over-represented KEGG biological pathways. The genes (rows) and samples (columns) were clustered using dendrograms built with Ward distance and hierarchical clustering. Anciently and recently invaded sites are indicated using respectively (AI) and (RI).

Figure 4. Protein-protein interaction network drawn from the 73 immune genes found to belong to enriched biological processes related to immunity. Nodes correspond to proteins, the thickness of edge network indicates the strength of data support, and the minimum required interaction score was set to 0.7 (high confidence level). The plain circle indicates the alipoprotein A – aloa - and haptoglobin – Hp – protein. The dashed circle indicates the complement proteins (C8). Fibrinogen (Fg) and serine peptidase inhibitor (serpin) proteins are distributed within the highly connected central part of the network.

References

1 Kamenova, S. *et al.* in *Networks of Invasion: A Synthesis of Concepts* Vol. 56 *Advances in Ecological Research* (eds D. A. Bohan, A. J. Dumbrell, & F. Massol) 85-182 (2017).

2 Kumschick, S. *et al.* Ecological impacts of alien species: quantification, scope, caveats, and recommendations. *Bioscience* **65**, 55-63 (2014).

3 Daszak, P., Cunningham, A. A. & Hyatt, A. D. Emerging infectious diseases of wildlife - threats to biodiversity and human health. *Science* **287**, 443-449, doi:10.1126/science.287.5452.443 (2000).

4 Bradshaw, C. J. A. *et al.* Massive yet grossly underestimated global costs of invasive insects. *Nature Communications* **7**, doi:10.1038/ncomms12986 (2016).

5 Facon, B. *et al.* A general eco-evolutionary framework for understanding bioinvasions. *Trends in Ecology & Evolution* **21**, 130-135, doi:10.1016/j.tree.2005.10.012 (2006).

6 Hufbauer, R. A. *et al.* Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evolutionary Applications* **5**, 89-101, doi:10.1111/j.1752-4571.2011.00211.x (2012).

7 Colautti, R. I. & Lau, J. A. Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology* **24**, 1999-2017, doi:10.1111/mec.13162 (2015).

8 Chuang, A. & Peterson, C. R. Expanding population edges: theories, traits, and trade-offs. *Global Change Biology* **22**, 494-512, doi:10.1111/gcb.13107 (2016).

9 Merrick, M. J. & Koprowski, J. L. Altered natal dispersal at the range periphery: The role of behavior, resources, and maternal condition. *Ecology & Evolution* **7**, 58-72, doi:10.1002/ece3.2612 (2017).

10 Torchin, M. E., Lafferty, K. D., Dobson, A. P., McKenzie, V. J. & Kuris, A. M. Introduced species and their missing parasites. *Nature* **421**, 628-630 (2003).

11 Torchin, M. E. & Mitchell, C. E. Parasites, pathogens, and invasions by plants and animals. *Frontiers in Ecology and the Environment* **2**, 183-190, doi:10.2307/3868313 (2004).

12 Dunn, A. M. in *Advances in Parasitology, Vol 68: Natural History of Host-Parasite Interactions* Vol. 68 *Advances in Parasitology* (ed J. P. Webster) 161-184 (2009).

13 Blossey, B. & Notzold, R. Evolution of increased competitive ability in invasive nonindigenous plants - a hypothesis. *Journal of Ecology* **83**, 887-889 (1995).

14 Lee, K. A. & Klasing, K. C. A role for immunology in invasion biology. *Trends in Ecology & Evolution* **19**, 523-529 (2004).

15 Cornet, S., Brouat, C., Diagne, C. A. & Charbonnel, N. EcoImmunology and bioinvasion: revisiting the EICA hypotheses. *Evolutionary Applications* **9**, 952–962 (2016).

16 Bernardi, G., Azzurro, E., Golani, D. & Miller, M. R. Genomic signatures of rapid adaptive evolution in the bluespotted cornetfish, a Mediterranean Lessepsian invader. *Molecular Ecology* **25**, 3384-3396, doi:10.1111/mec.13682 (2016).

17 Vera, M., Diez-del-Molino, D. & Garcia-Marin, J. L. Genomic survey provides insights into the evolutionary changes that occurred during European expansion of the invasive mosquitofish (*Gambusia holbrooki*). *Molecular Ecology* **25**, 1089-1105, doi:10.1111/mec.13545 (2016).

18 Hodgins, K. A., Lai, Z., Nurkowski, K., Huang, J. & Rieseberg, L. H. The molecular basis of invasiveness: differences in gene expression of native and introduced common ragweed (*Ambrosia artemisiifolia*) in stressful and benign environments. *Molecular Ecology* **22**, 2496-2510, doi:10.1111/mec.12179 (2013).

19 White, T. A., Perkins, S. E., Heckel, G. & Searle, J. B. Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology* **22**, 2971-2985, doi:10.1111/mec.12343 (2013).

20 Todd, E. V., Black, M. A. & Gemmell, N. J. The power and promise of RNA-seq in ecology and evolution. *Molecular Ecology* **25**, 1224-1241, doi:10.1111/mec.13526 (2016).

21 Alvarez, M., Schrey, A. W. & Richards, C. L. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Molecular Ecology* **24**, 710-725, doi:10.1111/mec.13055 (2015).

22 Rius, M. & Darling, J. A. How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution* **29**, 233-242, doi:10.1016/j.tree.2014.02.003 (2014).

23 Fraser, B. A., Ramnarine, I. W. & Neff, B. D. Temporal variation at the Mhc class IIB in wild populations of the guppy (*Poecilia reticulata*). *Evolution* **64**, 2086-2096 (2010).

24 Oleksiak, M. F., Churchill, G. A. & Crawford, D. L. Variation in gene expression within and among natural populations. *Nature Genetics* **32**, 261-266, doi:10.1038/ng983 (2002).

25 Whitehead, A. & Crawford, D. L. Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology* **15**, 1197-1211, doi:10.1111/j.1365-294X.2006.02868.x (2006).

26 Whitehead, A. & Crawford, D. L. Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5425-5430, doi:10.1073/pnas.0507648103 (2006).

27 Rollins, L. A., Richardson, M. F. & Shine, R. A genetic perspective on rapid evolution in cane toads (*Rhinella marina*). *Molecular Ecology* **24**, 2264-2276, doi:10.1111/mec.13184 (2015).

28 Vogel, H. J., Schmidtberg, H. & Vilcinskas, A. Comparative transcriptomics in three ladybird species supports a role for immunity in invasion biology. *Developmental & Comparative Immunology* **67**, 452-456 (2017).

29 Dalecky, A. *et al.* Range expansion of the invasive house mouse *Mus musculus domesticus* in Senegal, West Africa: West Africa: a synthesis of trapping data over three decades, 1983–2014. *Mammal Review* **45**, 176–190 (2015).

30 Konecny, A. *et al.* Invasion genetics of the introduced black rat (*Rattus rattus*) in Senegal, West Africa. *Molecular Ecology* **22**, 286-300, doi:10.1111/mec.12112 (2013).

31 Diagne, C. A. *et al.* Contemporary variations of immune responsiveness during range expansion of two invasive rodents in Senegal. *Oikos* **126**, 435-446 (2017).

32 Diagne, C. A. *et al.* Ecological and sanitary impacts of bacterial communities associated to biological invasions in African commensal rodent communities. *Nature Scientific Reports* **7**, 14995, doi:doi:10.1038/s41598-017-14880-1 (2017).

33 Diagne, C. A. *et al.* Parasites and invasions: changes in gastrointestinal helminth assemblages in invasive and native rodents in Senegal. *International Journal for Parasitology* **46**, 857–869 (2016).

34 Chuang, A. & Peterson, C. R. Expanding population edges: theories, traits, and trade-offs. *Global Change Biology* **22**, 494–512, doi:10.1111/gcb.13107 (2016).

35 DeBiasse, M. B. & Kelly, M. W. Plastic and evolved responses to global change: what can we learn from comparative transcriptomics? *Journal of Heredity* **107**, 71-81, doi:10.1093/jhered/esv073 (2016).

36 Lippens, C. *et al.* Genetic structure and invasion history of the house mouse (*Mus musculus domesticus*) in Senegal, West Africa: a legacy of colonial and contemporary times. *Heredity* **119**, 64-75, doi:10.1038/hdy.2017.18 (2017).

37 Whitehead, A., Triant, D. A., Champlin, D. & Nacci, D. Comparative transcriptomics implicates mechanisms of evolved pollution tolerance in a killifish population. *Molecular Ecology* **19**, 5186-5203, doi:10.1111/j.1365-294X.2010.04829.x (2010).

38 Barrett, S. C. H. Foundations of invasion genetics: the Baker and Stebbins legacy. *Molecular Ecology* **24**, 1927-1941, doi:https://doi.org/10.1111/mec.13014 (2015).

39 Dlugosch, K. M. & Parker, I. M. Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks. *Ecology Letters* **11**, 701-709, doi:10.1111/j.1461-0248.2008.01181.x (2008).

40 Chevin, L. M. & Lande, R. Adaptation to marginal habitats by evolution of increased phenotypic plasticity. *Journal of Evolutionary Biology* **24**, 1462-1476, doi:10.1111/j.1420-9101.2011.02279.x (2011).

41 Robertson, s., Bradley, J. E. & MacColl, A. D. Measuring the immune system of the three-spined stickleback – investigating natural variation by quantifying immune expression in the laboratory and the wild. *Molecular Ecology Resources* **16**, 701-713, doi:10.1111/1755-0998.12497 (2016).

42 Tengholm, A. & Gylfe, E. cAMP signalling in insulin and glucagon secretion. *Diabetes, Obesity & Metabolism* **19**, 42-53, doi:10.1111/dom .12993 (2017).

43 Jones, B. J., Tan, T. & Bloom, S. R. Minireview: Glucagon in Stress and Energy Homeostasis. *Endocrinology* **153**, 1049-1054, doi:10.1210/en.2011-1979 (2012).

44 Sih, A., Cote, J., Evans, M. R., Fogarty, S. & Pruitt, J. Ecological implications of behavioural syndromes. *Ecology Letters* **15**, 278-289, doi:10.1111/j.1461-0248.2011.01731.x (2012).

45 Bengston, S. E. *et al.* Genomic tools for behavioural ecologists to understand repeatable individual differences in behaviour. *Nature Ecology & Evolution* **2**, 944–955, doi:https://doi.org/10.1038/s41559-017-0411-4 (2018).

46 Schoener, T. W. The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science* **331**, 426-429, doi:10.1126/science.1193954 (2011).

47 Sikes, R. S. & Gannon, W. L. Guidelines of the American Society of Mammalogists for the use of wild mammals in research. *Journal of Mammalogy* **92**, 235-253 (2011).

48 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36 (2013).

49 Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359 (2012).

50 Anders, s., Pyl, T. P. & Huber, W. HTSeq — A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:doi: 10.1101/002824 (2015).

51 Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**, R80 (2004).

52 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

53 Simon, A. & Huber, W. Differential expression analysis for sequence count data. *Genome Biology* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).

54 R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria : the R Foundation for Statistical Computing (2011).

55 Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research* **45**, D362-D368, doi:10.1093/nar/gkw937 (2017).

56 Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43**, D447-452, doi:10.1093/nar/gku1003 (2015).

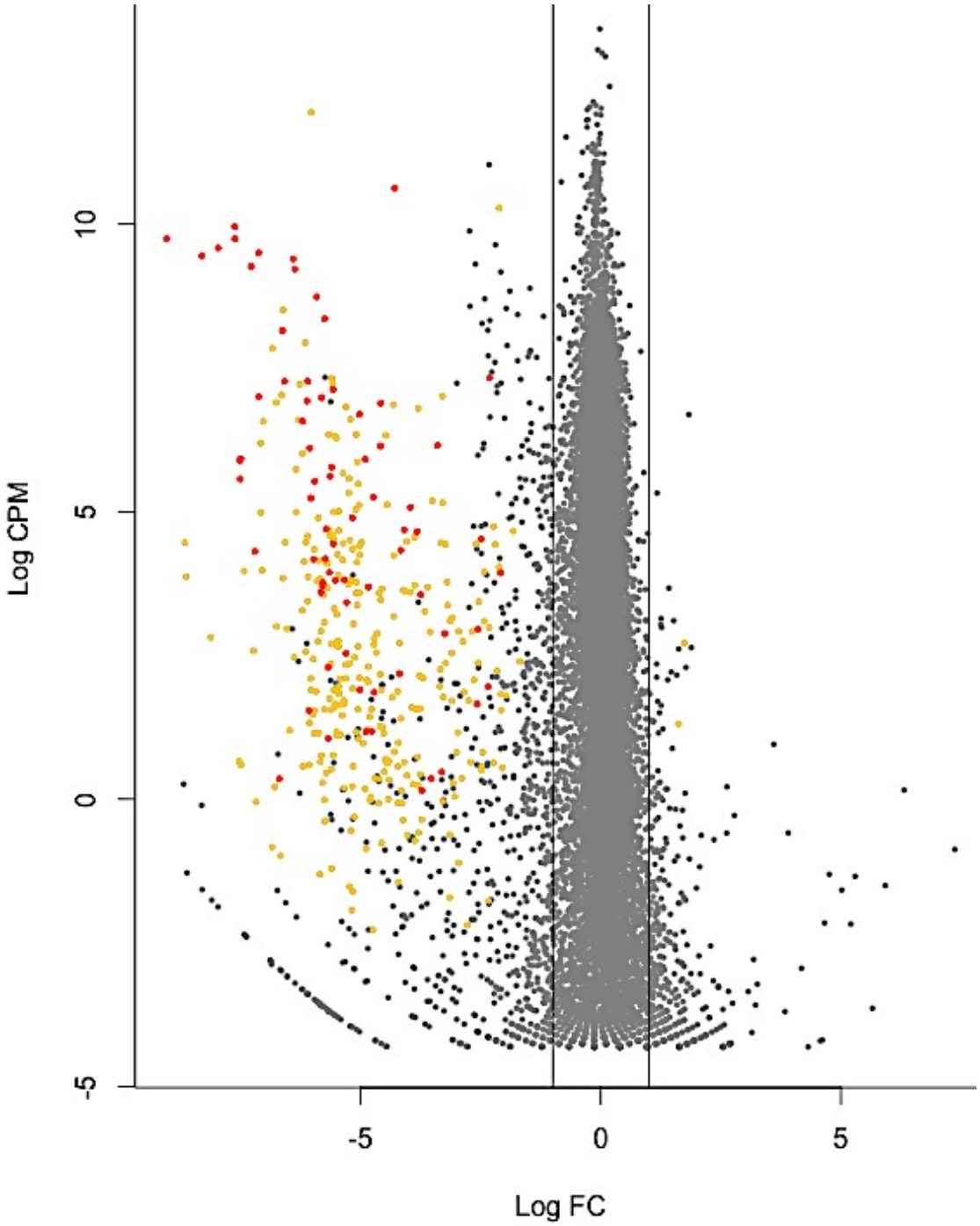
57 Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800, doi: https://doi.org/10.1371/journal.pone.0021800 (2011).

Table 1: Sampling sites and their invasion-related categorisation as well as rodent sample sizes included in each biological replicate on the house mouse and black rat invasion routes. Legend: AI: anciently invaded sites; RI: recently invaded sites; N: total number of rodents included; M: number of males; F: number of females.

House mouse		Black rat	
Sites (Invasion related category)	N (M/F)	Sites (Invasion related category)	N (M/F)
Dagathie (AI)		Diakene Wolof (AI)	
Pool of RNA #1	10 (5/5)	Pool of RNA #1	6 (3/3)
Pool of RNA #2	8 (4/4)	Pool of RNA #2	8 (4/4)
Mbakhana (AI)		Diattacounda (AI)	
Pool of RNA #1	10 (5/5)	Pool of RNA #1	10 (5/5)
Pool of RNA #2	10 (5/5)	Pool of RNA #2	10 (5/5)
Thilene (AI)		Marsassoum (AI)	
Pool of RNA #1	8 (4/4)	Pool of RNA #1	10 (5/5)
Pool of RNA #2	8 (4/4)	Pool of RNA #2	10 (5/5)
Ndombo (AI)		Tobor (AI)	
Pool of RNA #1	10 (5/5)	Pool of RNA #1	6 (3/3)
Pool of RNA #2	10 (5/5)	Pool of RNA #2	8 (4/4)
Croisement Boube (RI)		BadiNieriko (RI)	
Pool of RNA #1	10 (5/5)	Pool of RNA #1	8 (4/4)
Pool of RNA #2	10 (5/5)	Pool of RNA #2	8 (4/4)
Dodel (RI)		Boutougoufara (RI)	
Pool of RNA #1	10 (5/5)	Pool of RNA #1	8 (4/4)
Pool of RNA #2	8 (4/4)	Pool of RNA #2	8 (4/4)
Aere Lao (RI)		Kedougou (RI)	
Pool of RNA #1	10 (5/5)	Pool of RNA #1	8 (4/4)
Pool of RNA #2	10 (5/5)	Pool of RNA #2	8 (4/4)
Lougue (RI)		Soutouta (RI)	
Pool of RNA #1	10 (5/5)	Pool of RNA #1	10 (5/5)
Pool of RNA #2	8 (4/4)	Pool of RNA #2	10 (5/5)

Figure 1. Significantly differentially expressed genes (orange dots) and immune related genes (red dots) between anciently invaded and recently invaded sites along a) the house mouse and b) the black rat invasion routes. Only the 73 immune related genes belonging to biological pathways found to be significantly enriched are represented for the house mouse. Vertical lines indicate 1 log fold change (Log FC). The x-axis indicates genes that are down- (negative values) and up-regulated (positive values) in anciently compared to recently invaded areas. The y-axis represents the average level of gene expression using read counts per million (Log CPM).

a)



b)

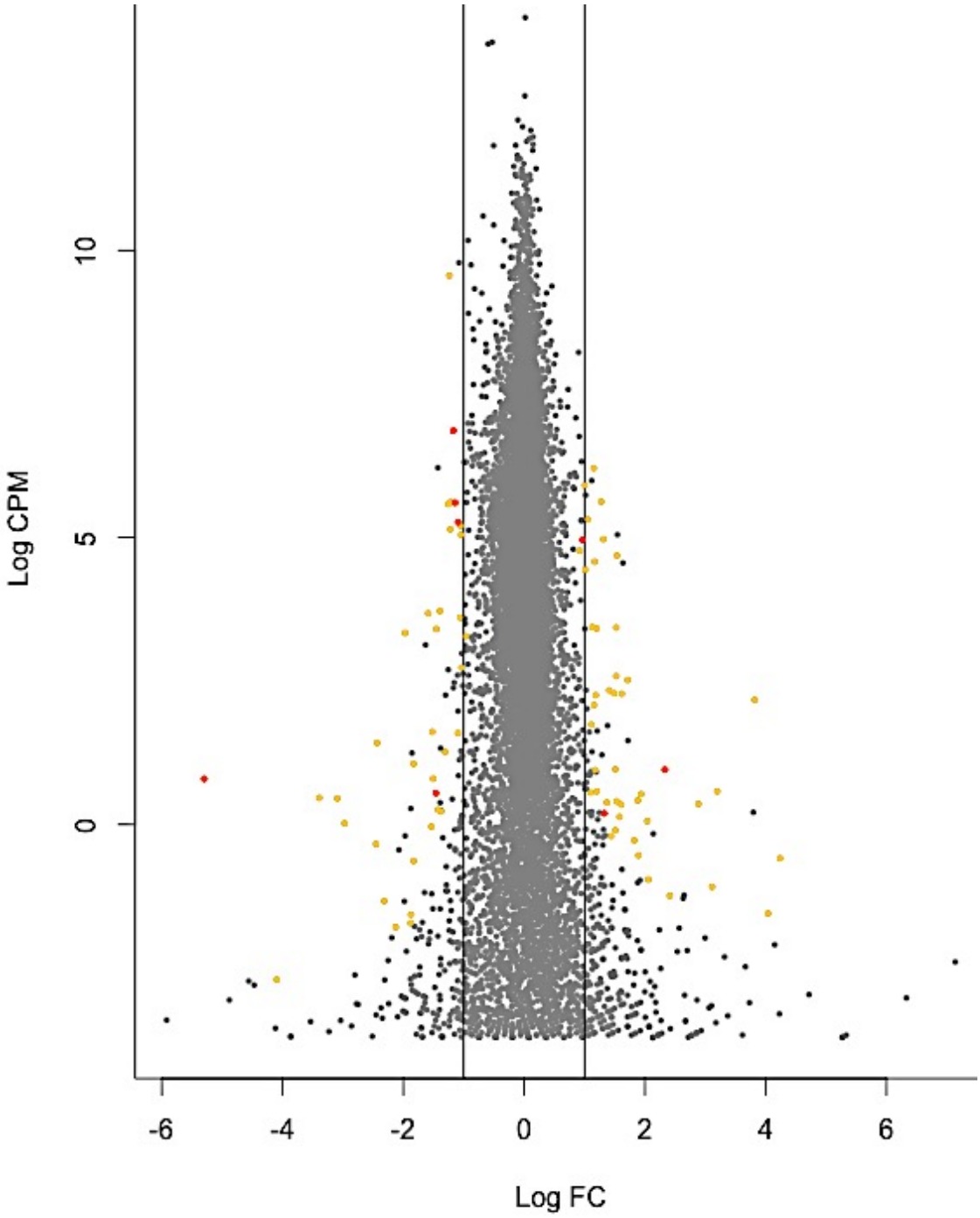


Figure 2. TreeMap view of REVIGO Biological Process analyses for the house mouse differentially expressed (DE) robust genes. Each rectangle represents a single cluster. Large groups represent 'superclusters' of related gene ontology (GO) terms, and they are represented with different colours. The size of the rectangles reflects the frequency of the GO term in this set of DE genes.

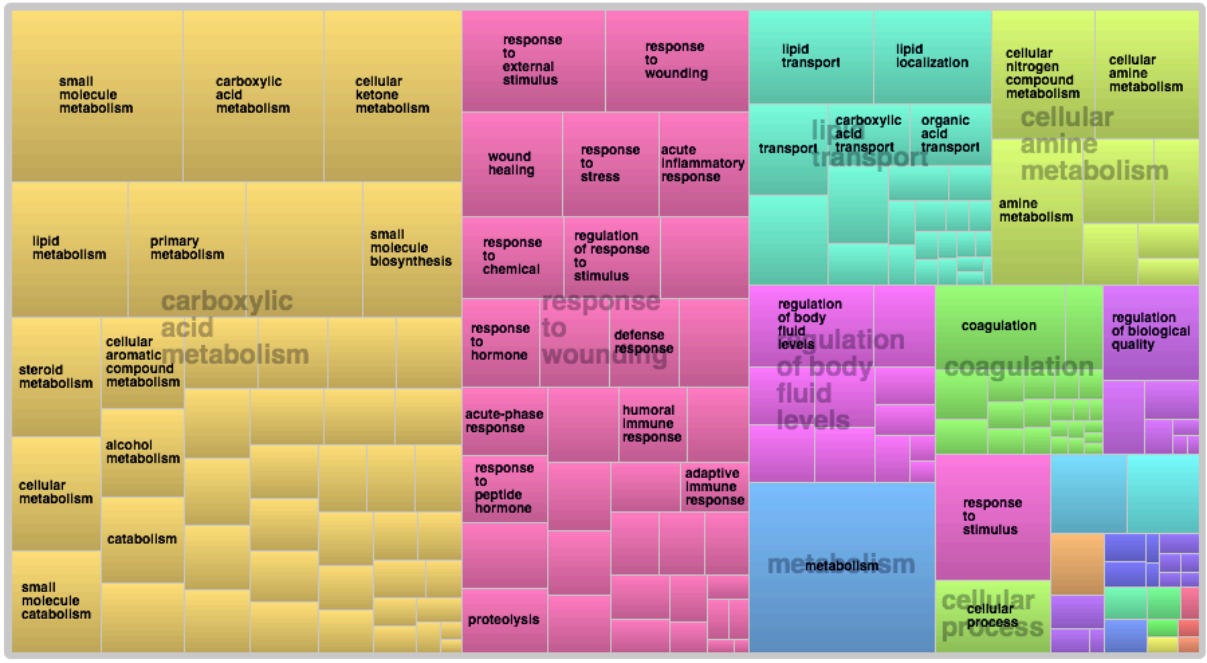


Figure 3. Heatmap of the differentially expressed (DE) genes between the anciently and recently invaded sites of house mouse (*M. musculus domesticus*). The normalized read counts for the expressed genes are shown. For clarity, the heatmap was built in R using heatmap.2 for 29 immune related genes belonging to over-represented KEGG biological pathways. The genes (rows) and samples (columns) were clustered using dendrograms built with Ward distance and hierarchical clustering. Anciently and recently invaded sites are indicated using respectively (AI) and (RI).

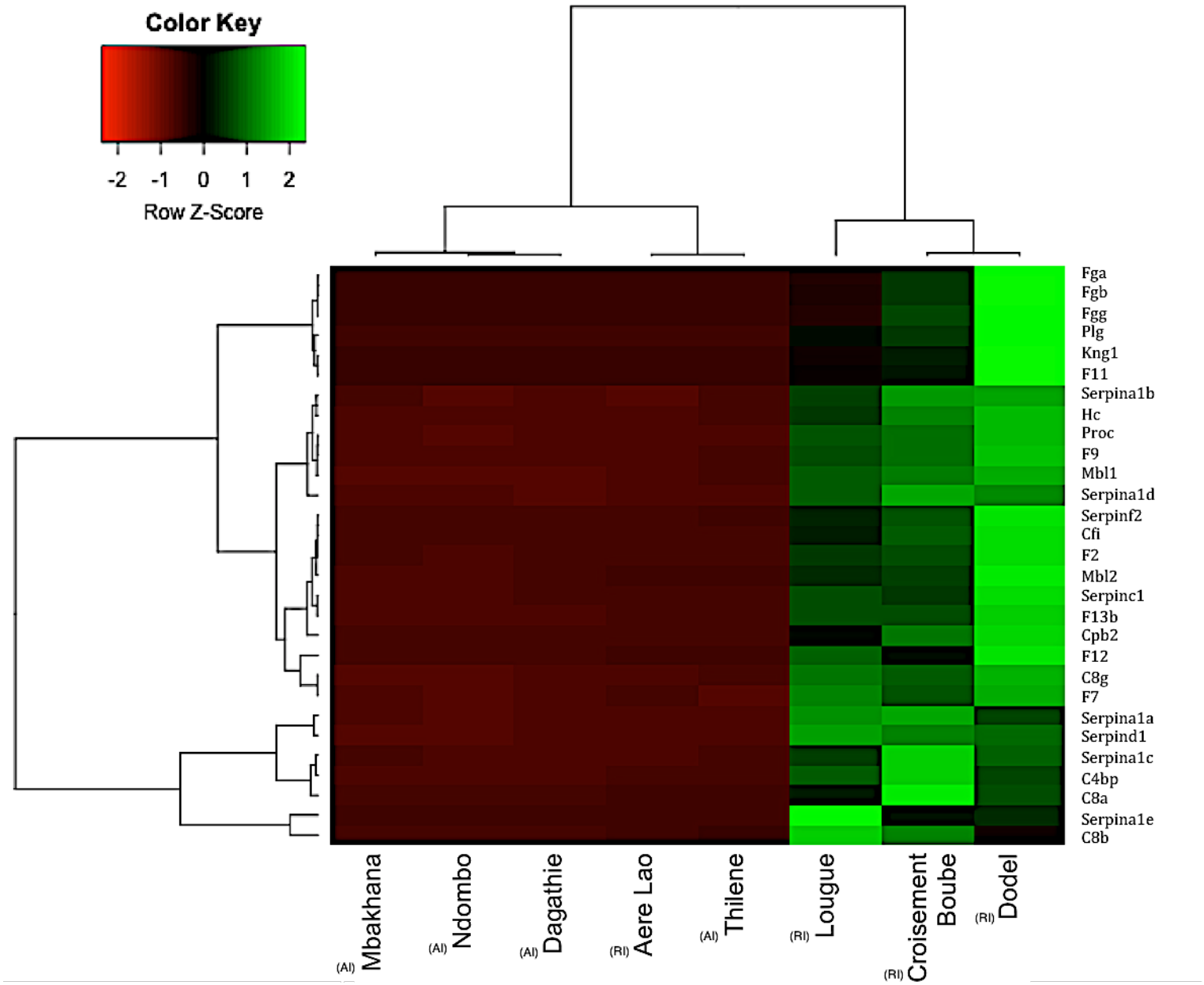

Color Key
-2
-1
0
1
2
Row Z-Score
Fga
Fgb
Fgg
Plg
Kng1
F11
Serpina1b
Hc
Proc
F9
Mbl1
Serpina1d
Serpinf2
Cfi
F2
Mbl2
Serpinc1
F13b
Cpb2
F12
C8g
F7
Serpina1a
Serpind1
Serpina1c
C4bp
C8a
Serpina1e
C8b
(AI) Mbakhana
(AI) Ndombo
(AI) Dagathie
(RI) Aere Lao
(AI) Thilene
(RI) Lougue
(RI) Croisement Boube
(RI) Dodel

Figure 4. Protein-protein interaction network drawn from the 73 immune genes found to belong to enriched biological processes related to immunity. Nodes correspond to proteins, the thickness of edge network indicates the strength of data support, and the minimum required interaction score was set to 0.7 (high confidence level). The plain circle indicates the alipoprotein A – aloa - and haptoglobin – Hp – protein. The dashed circle indicates the complement proteins (C8). Fibrinogen (Fg) and serine peptidase inhibitor (serpin) proteins are distributed within the highly connected central part of the network.

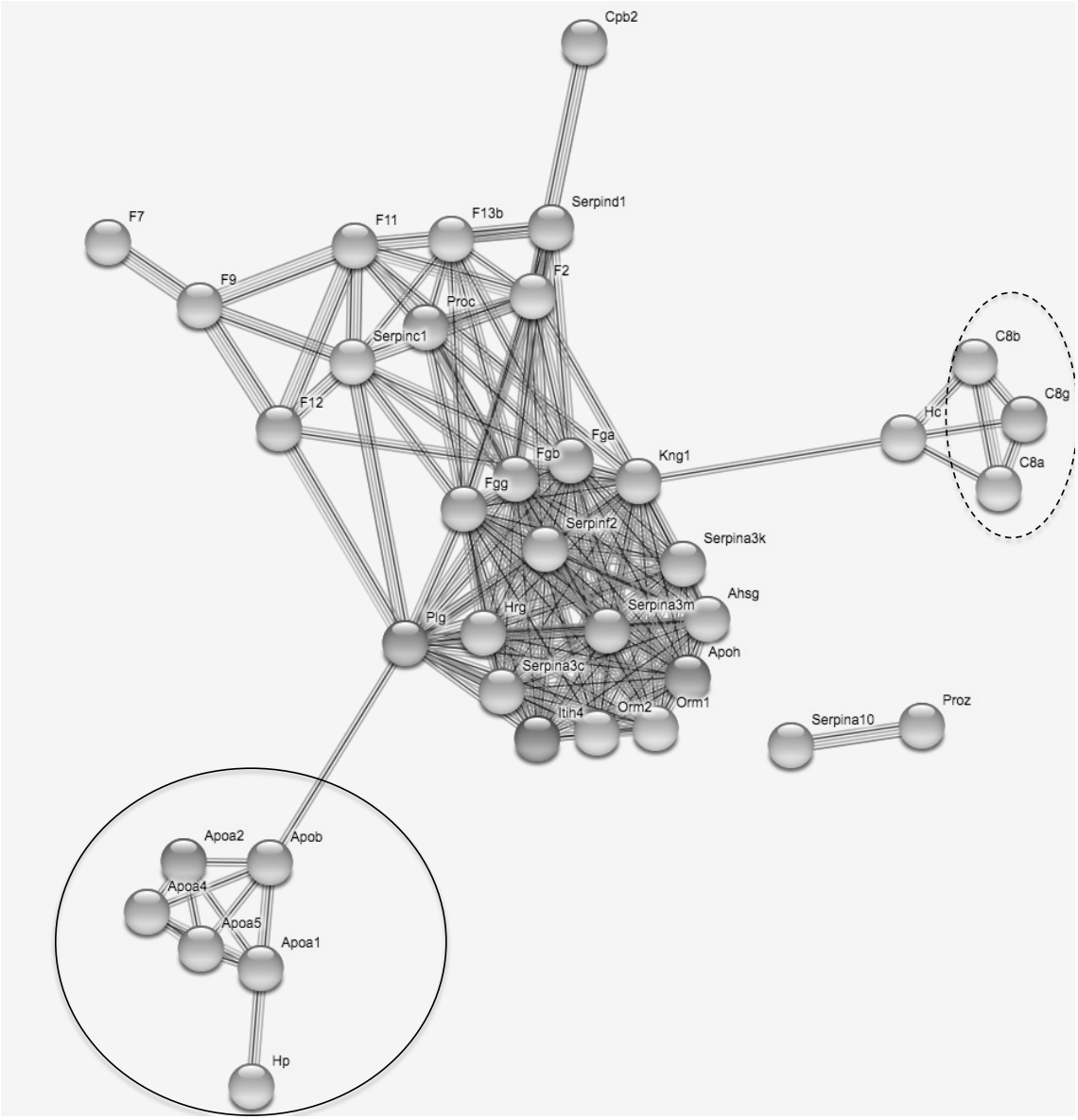
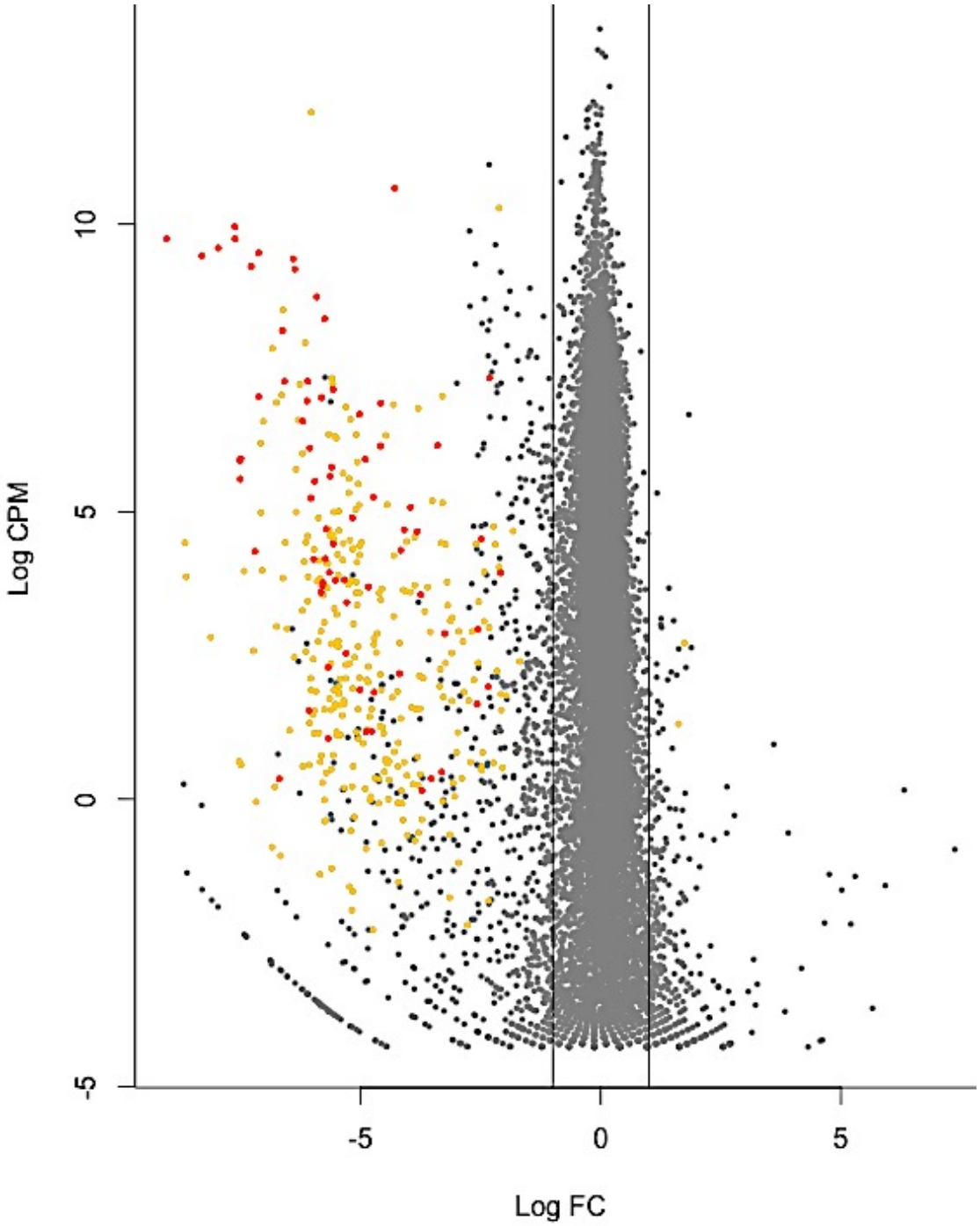


Figure 1. Significantly differentially expressed genes (orange dots) and immune related genes (red dots) between anciently invaded and recently invaded sites along a) the house mouse and b) the black rat invasion routes. Only the 73 immune related genes belonging to biological pathways found to be significantly enriched are represented for the house mouse. Vertical lines indicate 1 log fold change (Log FC). The x-axis indicates genes that are down- (negative values) and up-regulated (positive values) in anciently compared to recently invaded areas. The y-axis represents the average level of gene expression using read counts per million (Log CPM).

a)



b)

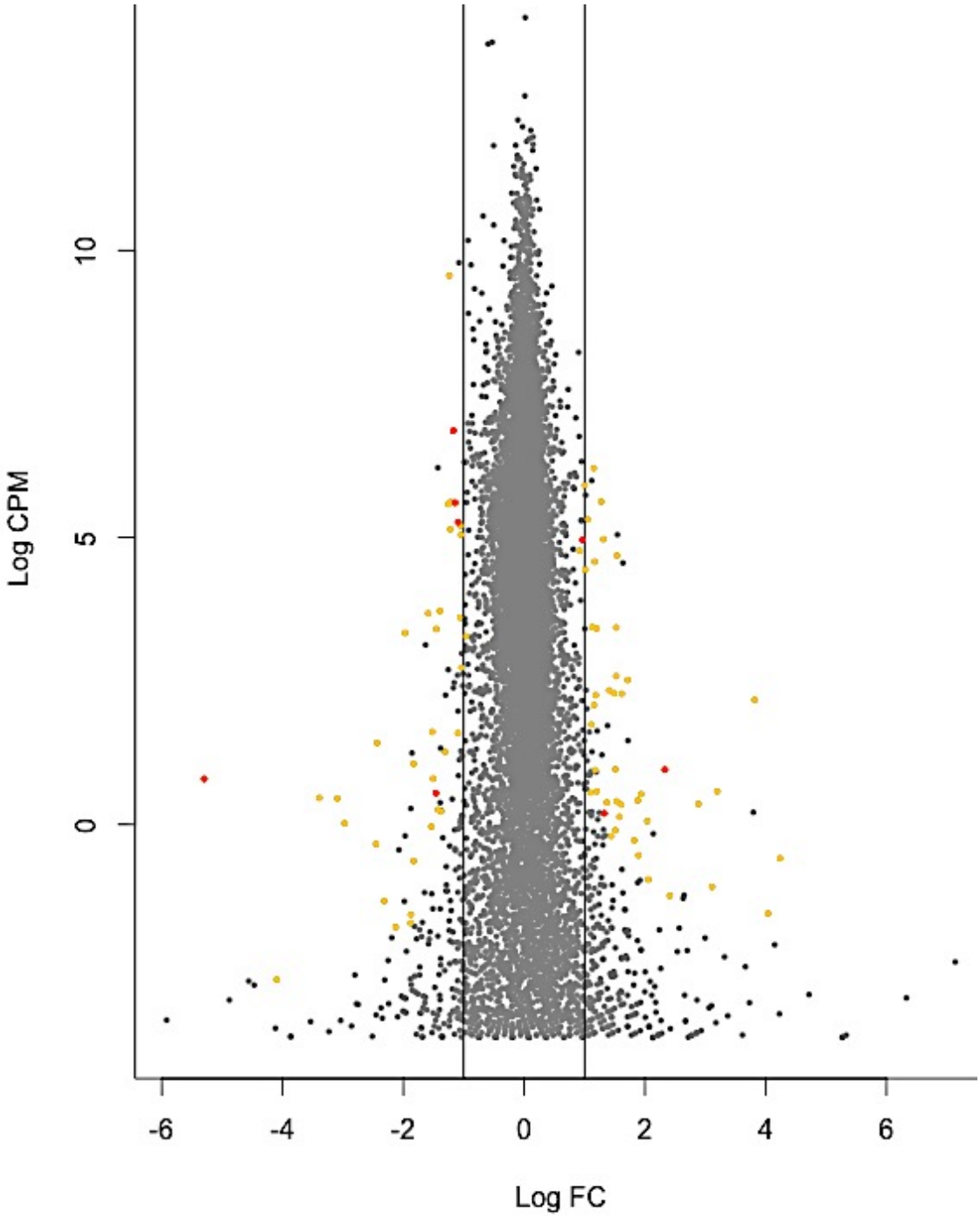


Figure 2. TreeMap view of REVIGO Biological Process analyses for the house mouse differentially expressed (DE) robust genes. Each rectangle represents a single cluster. Large groups represent 'superclusters' of related gene ontology (GO) terms, and they are represented with different colours. The size of the rectangles reflects the frequency of the GO term in this set of DE genes.

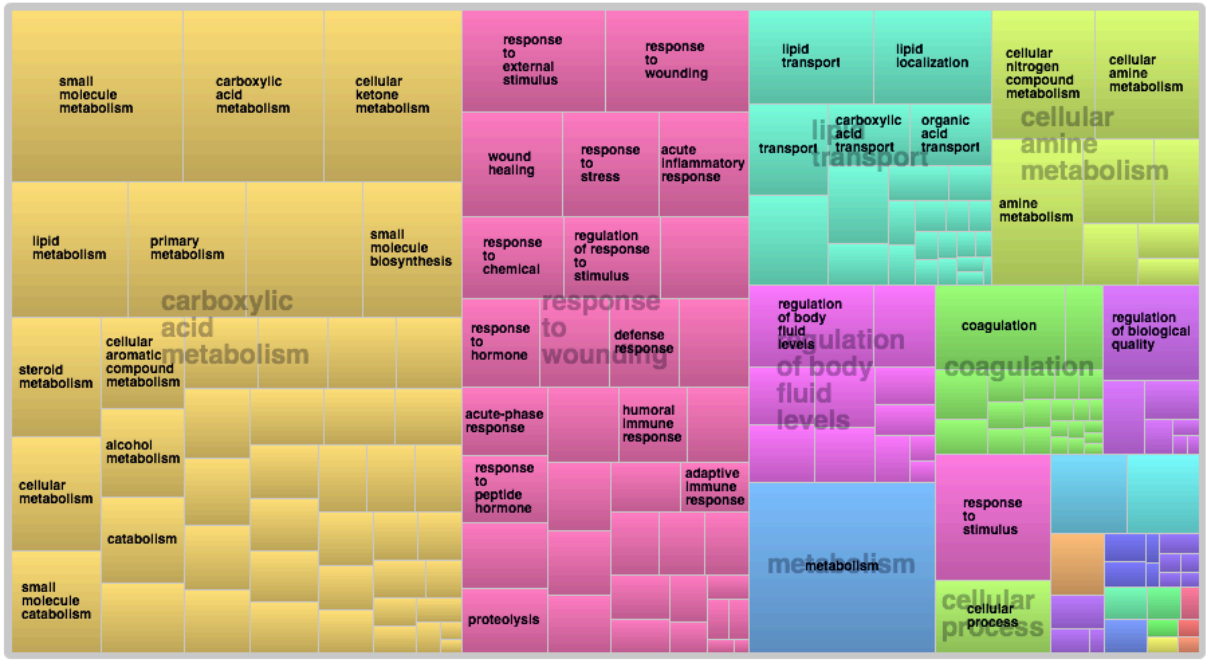
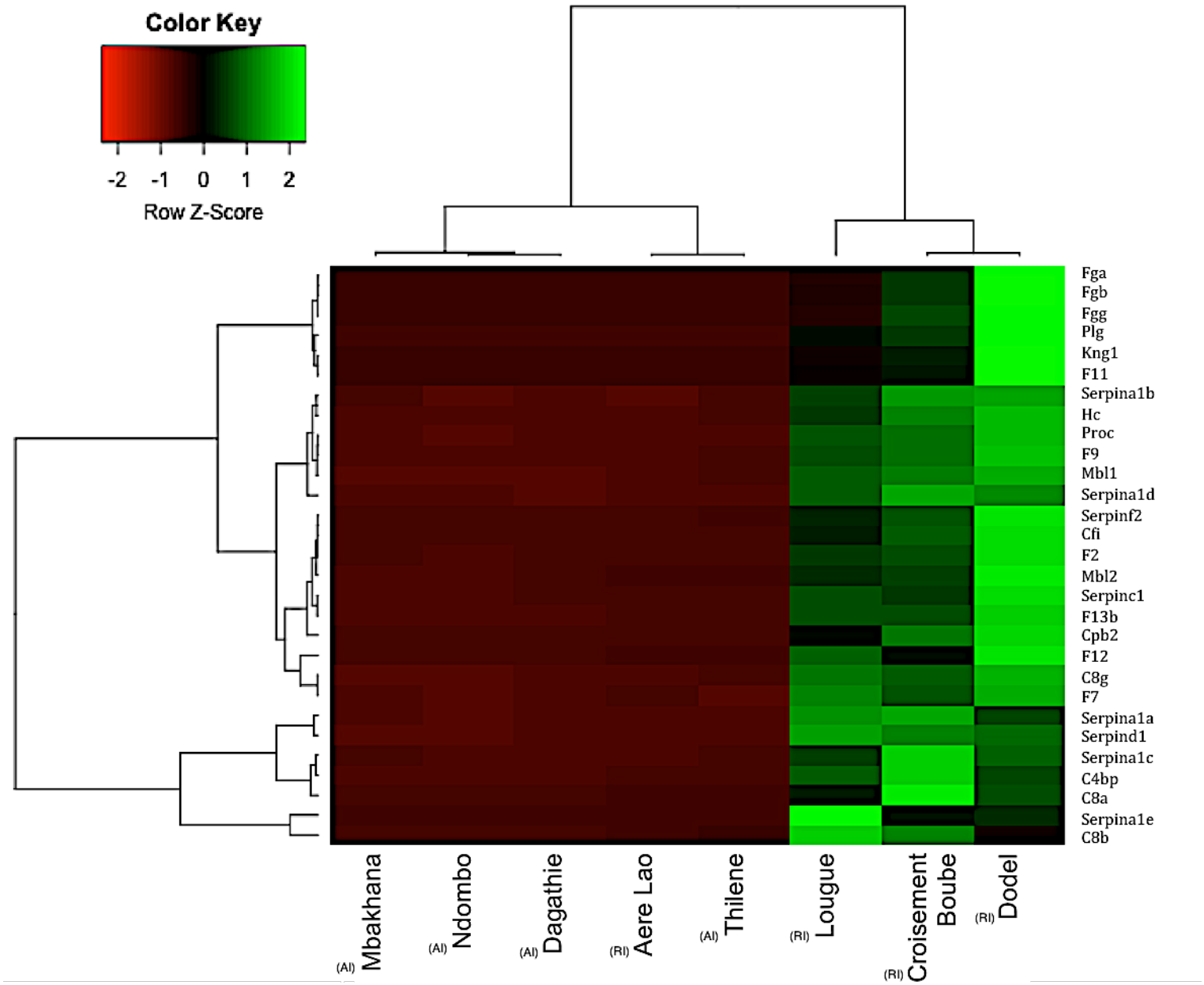


Figure 3. Heatmap of the differentially expressed (DE) genes between the anciently and recently invaded sites of house mouse (*M. musculus domesticus*). The normalized read counts for the expressed genes are shown. For clarity, the heatmap was built in R using heatmap.2 for 29 immune related genes belonging to over-represented KEGG biological pathways. The genes (rows) and samples (columns) were clustered using dendrograms built with Ward distance and hierarchical clustering. Anciently and recently invaded sites are indicated using respectively (AI) and (RI).


Color Key
-2
-1
0
1
2
Row Z-Score
Fga
Fgb
Fgg
Plg
Kng1
F11
Serpina1b
Hc
Proc
F9
Mbl1
Serpina1d
Serpinf2
Cfi
F2
Mbl2
Serpinc1
F13b
Cpb2
F12
C8g
F7
Serpina1a
Serpind1
Serpina1c
C4bp
C8a
Serpina1e
C8b
(AI) Mbakhana
(AI) Ndombo
(AI) Dagathie
(RI) Aere Lao
(AI) Thilene
(RI) Lougue
(RI) Croisement Boube
(RI) Dodel

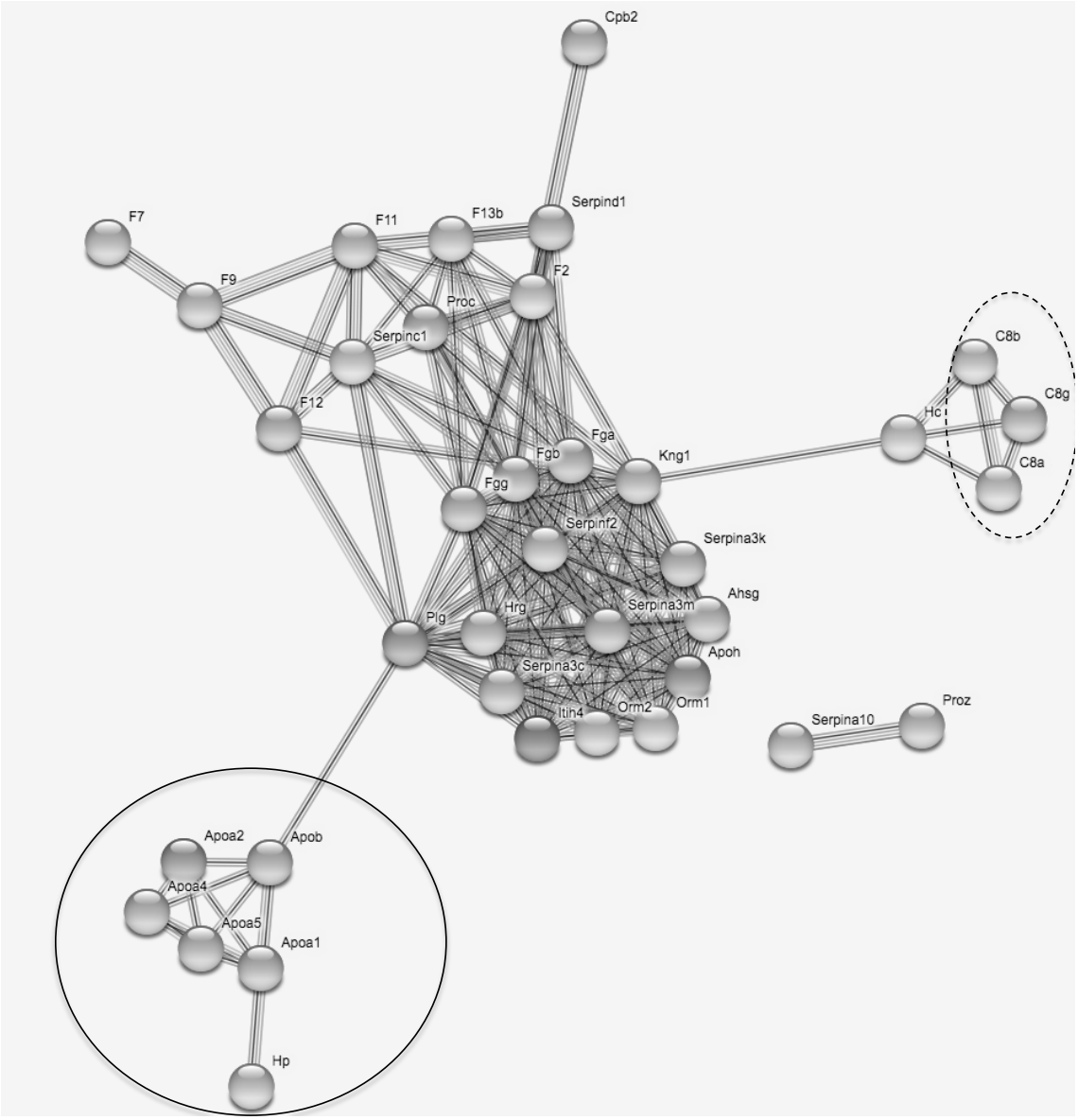



Figure 4. Protein-protein interaction network drawn from the 73 immune genes found to belong to enriched biological processes related to immunity. Nodes correspond to proteins, the thickness of edge network indicates the strength of data support, and the minimum required interaction score was set to 0.7 (high confidence level). The plain circle indicates the alipoprotein A – aloa - and haptoglobin – Hp – protein. The dashed circle indicates the complement proteins (C8). Fibrinogen (Fg) and serine peptidase inhibitor (serpin) proteins are distributed within the highly connected central part of the network.