



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)



Infection, Genetics and Evolution 00 (2015) 1–10

---

---

Infection, Genetics  
and Evolution

---

---

## Inducible Nitric Oxide Synthase (*iNOS*) Regulatory Region Variation in Non-Human Primates

Morteza Roodgar<sup>2,3,4\*</sup>, Cody T. Ross<sup>3,4,5</sup>, Nicholas J. Kenyon<sup>1</sup>, Gretchen Marcelino<sup>3</sup>, David Glenn Smith<sup>3,4</sup>

<sup>1</sup> University of California, Davis School of Medicine, Department of Internal Medicine, Division of Pulmonary and Critical Care

<sup>2</sup> Department of Veterinary Medicine, University of California, Davis

<sup>3</sup> Molecular Anthropology Laboratory, University of California, Davis

<sup>4</sup> Department of Anthropology, University of California, Davis

<sup>5</sup> Santa Fe Institute, Santa Fe, New Mexico

---

### Abstract

Inducible nitric oxide synthase (*iNOS*) is an enzyme that plays a key role in intracellular immune response against respiratory infections. Since various species of nonhuman primates exhibit different levels of susceptibility to infectious respiratory diseases, and since variation in regulatory regions of genes is thought to play a key role in expression levels of genes, two candidate regulatory regions of *iNOS* were mapped, sequenced, and compared across five species of nonhuman primates: African green monkeys (*Chlorocebus sabaeus*), pig-tailed macaques (*Macaca nemestrina*), cynomolgus macaques (*Macaca fascicularis*), Indian rhesus macaques (*Macaca mulatta*), and Chinese rhesus macaques (*Macaca mulatta*). In addition, we conducted an *in silico* analysis of the transcription factor binding sites associated with genetic variation in these two candidate regulatory regions across species. We found that only one of the two candidate regions showed strong evidence of involvement in *iNOS* regulation. Specifically, we found evidence of 13 conserved binding site candidates linked to *iNOS* regulation: AP-1, C/EBPB, CREB, GATA-1, GATA-3, NF-AT, NF-AT5, NF- $\kappa$ B, KLF4, Oct-1, PEA3, SMAD3, and TCF11. Additionally, we found evidence of interspecies variation in binding sites for several regulatory elements linked to *iNOS* (GATA-3, GATA-4, KLF6, SRF, STAT-1, STAT-3, OLF-1 and HIF-1) across species, especially in African green monkeys relative to other species. Given the key role of *iNOS* in respiratory immune response, the findings of this study might help guide the direction of future studies aimed to uncover the molecular mechanisms underlying the increased susceptibility of African green monkeys to several viral and bacterial respiratory infections.

© 2011 Elsevier Ltd. All rights reserved

**Keywords:** *iNOS*, *NOS2A*, Immune Response, Non-Human Primates, Regulatory Region

---

---

\*Corresponding authors

Email addresses: [mroodgar@ucdavis.edu](mailto:mroodgar@ucdavis.edu) (Morteza Roodgar<sup>2,3,4</sup>), [ctross@ucdavis.edu](mailto:ctross@ucdavis.edu) (Cody T. Ross<sup>3,4,5</sup>)

1    **1. Introduction**

2       Inducible nitric oxide synthase (*iNOS*) is an enzyme that plays a key role in immune response against pathogens  
 3 through the production of peroxinitrite in macrophages [1, 2]. Genetic changes in the regulatory and/or coding re-  
 4 gions of *NOS2A*, the gene encoding *iNOS*, might play an important role in modulating expression levels of *iNOS*  
 5 and, consequently, cause variation in immune response against intracellular pathogens [3]. Several studies indicate  
 6 that genetic changes in the regulatory and/or coding region of *NOS2A* are associated with susceptibility to various  
 7 diseases [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. It has been shown, for example, that mutations in the promoter region of  
 8 *NOS2A* correlate with susceptibility to malaria [13] and that *iNOS* expression levels are strongly driven by exposure  
 9 to *Mycobacterium tuberculosis* vaccination [14]. In addition, recent epigenetic studies of the promoter and enhancers  
 10 of *NOS2A* in humans demonstrate the effect of changes in *NOS2A* regulatory regions on *iNOS* expression [15] and  
 11 the pathogenesis of infectious diseases [16, 17, 18, 19].

12      In this study, we investigated the patterning and functional significance of variation in two candidate regulatory  
 13 regions of *NOS2A* (which we label *R1* and *R2* for notational convenience, see definitions in *Material and Methods*)  
 14 across five taxa of non-human primates (NHPs) relevant to biomedical research: African green monkeys (*Chlorocebus*  
 15 *sabaeus*), pig-tailed macaques (*Macaca nemestrina*), cynomolgus macaques (*Macaca fascicularis*), Indian rhesus  
 16 macaques (*Macaca mulatta*), and Chinese rhesus macaques (*Macaca mulatta*). Variation in regulation of *iNOS*  
 17 expression may play a role in the variable susceptibility of these species to several infectious diseases [14, 20, 21]. We  
 18 investigated whether or not there is evidence for interspecies differences in the regulatory regions of *iNOS* that might  
 19 account for such variability in disease susceptibility. Since *NOS2A* plays a key role in immunity against intracellular  
 20 pathogens [22, 23] relevant to human health, information on DNA sequence variation in the regulatory region of  
 21 *NOS2A* in species of NHP that are more closely related to humans than the mouse should provide better information  
 22 about the relationship between variation in *NOS2A* gene expression and human-like immune responses to intracellular  
 23 pathogens [20, 21].

24      We sequenced two candidate regulatory regions of the *NOS2A* gene in several animals in each of five species  
 25 or subspecies of NHP that exhibit differing levels of susceptibility to infectious respiratory diseases, especially tu-  
 26 berculosis. The basic primer sequences for the candidate *iNOS* promoter regions were identified using the human  
 27 genome Chip-seq data available at the University of California, Santa Cruz (UCSC) Genome Browser. We then used  
 28 an Applied Biosystems 3130XL genetic analyzer to produce DNA sequence data for each sample. Sequences were  
 29 aligned using Kalign2 [24, 25, 26], and variation in the candidate promoter regions was analyzed using the *adegenet*  
 30 package in the R programming environment [27]. The effect of cross-species genetic conservation and variation on  
 31 transcription factor and regulatory element bindings sites was then evaluated using the MatInspector software [28, 29].

32    **2. Results**

33    **2.1. Multiple Sequence Alignment and Promoter Localization**

34      We localized the *iNOS* coding region and the candidate promoter/regulatory regions on the rhesus macaque and  
 35 human reference sequences using the NCBI genome browser. Figure 1 plots the location of the *R2* region on the  
 36 human reference sequence.

37      [Figure 1 about here.]

38      Multiple sequence alignment of *R1* and *R2* across species was conducted using Kalign2 [24, 25, 26]. The perfor-  
 39 mance of the alignment was evaluated using Mumsa [25] and visual inspection. The Jalview program [30, 31] was  
 40 used to visualize and trim the alignments and construct species-specific consensus sequences. In Figure 2, we plot the  
 41 animal-specific nucleotide sequences and species-specific consensus sequences used in this analysis.

42      [Figure 2 about here.]

## 2.2. Promoter Variation and Interspecies Clustering

To investigate whether or not variation in *R1* and *R2* followed the pattern expected from the phylogenetic relationships among these NHP species and subspecies, we used the R packages *ape* and *adegenet* to extract the SNPs from the aligned and trimmed DNA sequences. We then used discriminant analysis of principal components (DAPC) [32] to investigate the cross-species partitioning of genetic variation.

Figure 3 plots the location of all animals included in this study on the first two principal components of variation. Figure 3 illustrates that for both *R1* and *R2* the first principal component separates African green monkeys from the other species. The second principal component separates Chinese and Indian rhesus macaques from cynomolgus and pig-tailed macaques. In the *R1* region, the second principal component fails to separate Chinese rhesus macaques from Indian rhesus macaques and cynomolgus macaques from pig-tailed macaques. In the *R2* region, the second principal component separates Chinese rhesus macaques from Indian rhesus macaques to some extent, while cynomolgus macaques and pig-tailed macaques remain unseparated.

[Figure 3 about here.]

Figure 4 illustrates the group assignment probabilities based on the DAPC analysis of SNPs in the *R1* and *R2* regions. We find that African green monkeys can be distinguished from other species with high probability at both *R1* and *R2*. In *R2*, but not *R1*, Indian rhesus, and to some extent Chinese rhesus, can be discriminated from other species with high probability. In both *R1* and *R2*, cynomolgus macaques cannot be distinguished from pig-tailed macaques.

[Figure 4 about here.]

## 2.3. In Silico Regulatory Element and Transcription Factor Binding Site Analysis

To understand the possible phenotypic consequence of genetic variation in the *R1* and *R2* regions and identify whether or not each region is likely to be involved in regulation of *iNOS* transcription, we conducted an *in silico* analysis of regulatory element (RE) and transcription factor (TF) binding to the DNA sequences in *R1* and *R2* using MatInspector [28, 29]. We identified several key RE and TF bindings sites inside both *R1* and *R2*. Many of these sites were conserved across all species included in this analysis. Notably, region *R2* contains binding sites for the majority of REs and TFs known from laboratory studies to influence *iNOS* expression (see *Discussion*), while the *R1* region lacks binding sites for almost all of these elements.

We also identified several species-specific RE and TF bindings sites, indicating that variation in TF and RE binding to the DNA sequences of *R1* and *R2* is influenced by the SNPs discovered in this analysis. We detail these findings in the subsections that follow.

### 2.3.1. Cross Species Conservation

We found 39 RE/TF binding sites that were conserved across all animals in *R1* and 71 that were conserved across all animals in *R2*. Supplementary Table 1 contains the full list of the MatInspector Matrix IDs of these REs and TFs. Notably, in *R2*, binding sites for *AP-1*, *C/EBPB*, *CREB*, *GATA-1*, *GATA-3*, *NF-AT*, *NF-AT5*, *NF- $\kappa$ B*, *KLF4*, *Oct-1*, *PEA3*, *SMAD3*, and *TCF11*, 13 genes that have been previously associated with *iNOS* regulation [33, 34], were conserved. In contrast, only 2 binding sites previously associated with *iNOS* regulation were conserved in *R1* (*PPAR- $\gamma$*  and *USF-1*) [33].

### 2.3.2. Cross Species Variation

We identified 105 RE/TF binding sites in *R1* and 95 RE/TF binding sites in *R2* which included a SNP that disrupted the simulated binding of the REs and TFs to the DNA sequence. Supplementary Table 2 contains the full list of the MatInspector Matrix IDs of these REs and TFs, paired with annotations and the animal specific data. To identify which of these binding sites were predictive of interspecies differences, we used an  $\ell_1$ -regularized categorical Bayesian multiple regression model to classify animals into three groups (African green monkeys, Chinese/Indian rhesus macaque, or cynomolgus/pig-tailed macaque), using the existence or non-existence of RE and TF bindings sites as predictors. The maximum *a posteriori* parameter vector representing the strength of association between any RE or TF binding site and all three outcome categories can be plotted in barycentric space using a tertiary plot to represent the relative strength of association across categories. Figure 5 displays such plots for *R1* and *R2*. Each colored letter represents an RE or TF binding site and links it to its gene name in Supplementary Table 3.

90 [Figure 5 about here.]

91 Most points in the tertiary plot are shrunken towards the center of the plotting space by the regularizing priors  
 92 of the Bayesian model, showing that most variation in binding sites is not indicative of cross species differences.  
 93 However, a few RE/TF binding sites in both *R1* and *R2* reflect differences between African green monkeys and the  
 94 other species. For example, in *R1*, binding sites for *GATA-3*, *MEIS1*, *PU.1/SPI-1*, and *EBF1* show increased  
 95 association with African green monkeys, and those for *HOXB8*, *GATA-4*, *DLX3*, *CDX1*, *HOXA9*, *HOXB6*, *IRF4*,  
 96 *ZNF384*, *TEF*, *NF-AT5*, *MSX3*, *MEOX1*, and *HSF2* show a decreased association with African green monkeys,  
 97 relative to other species.

98 In *R2* we find an increased association of binding sites for *GCM1*, *GLIS2*, *KLF6*, *MAZ*, *MTBF*, *NM23*, *PAX6*,  
 99 *RTTR*, *SRF*, *STAT-1*, *STAT-3*, *ZIC2*, and *ZNF219* with African green monkeys, and a decreased association of  
 100 *ELF-1*, *GLS3*, *HOXB9*, *S MARCA2*, *OLF-1*, and *TIEG* with African green monkeys, relative to other species.

101 It is also notable that binding sites for *PAX3* and *HIF-1* are negatively associated with rhesus macaques relative  
 102 to the other groups.

### 103 3. Discussion

#### 104 3.1. Promoter Variation and Interspecies Clustering

105 The results of this study indicate that genetic divergence in *R1* and *R2* across species roughly corresponds to  
 106 that which would be expected from phylogenetic relationships. African green monkeys exhibit the most distant  
 107 evolutionary relationship with the other species, followed in order by pig-tailed macaques, cynomolgus macaques,  
 108 Chinese rhesus macaques (with gene flow to cynomolgus macaques), and finally Indian rhesus macaques, the most  
 109 highly derived of the five taxa [35]. However, in contrast to expectations under purely neutral evolution, pig-tailed  
 110 and cynomolgus macaques exhibit greater clustering than cynomolgus and Chinese rhesus macaques even though  
 111 the evolutionary distance between pig-tailed and cynomolgus macaques is greater than that between cynomolgus and  
 112 Chinese rhesus macaques [36]. This pattern of divergence of rhesus from other species might be indicative of rhesus  
 113 specific selection or demographic effects.

#### 114 3.2. In Silico Regulatory Element and Transcription Factor Binding Site Analysis

##### 115 3.2.1. Conserved RE/TF Binding Sites

116 In this study, we discovered 13 conserved binding site candidates linked to *iNOS* regulation (*AP-1*, *C/EBPB*,  
 117 *CREB*, *GATA-1*, *GATA-3*, *NF-AT*, *NF-AT5*, *NF- $\kappa$ B*, *KLF4*, *Oct-1*, *PEA3*, *SMAD3*, and *TCF11*) in *R2* across  
 118 species. In addition to the results of *in silico* analyses presented here [33, 34, 37], many of these transcription factors  
 119 have been shown to play a key role in the regulation of *iNOS* expression in laboratory studies. In contrast, we  
 120 identified only two conserved binding sites previously linked to *iNOS* regulation (*PPAR- $\gamma$*  and *USF-1*) in *R1* [33].  
 121 These results provide evidence that *R2* is a more appropriate candidate region for *iNOS* regulation in these species of  
 122 NHPs than *R1*.

123 In the list of conserved RE/TF binding sites across these species of NHP, *NF- $\kappa$ B* is the most well-studied transcrip-  
 124 tion factor known an important role in regulation of *iNOS* expression [38, 39, 40, 41, 42, 43]. Previous studies have  
 125 shown that *NF- $\kappa$ B* modulates *iNOS* expression during various viral, bacterial, and parasitic infections [41, 39, 40]. It  
 126 has also been shown that interference in the *NF- $\kappa$ B*-dependent regulation of *iNOS* expression is part of the pathogenic  
 127 mechanism for *Helicobacter pylori* [44]; a similar mechanism may be involved in the pathogenesis of other infectious  
 128 diseases (e.g., *Mycobacterium tuberculosis*).

##### 129 3.2.2. Non-Conserved RE/TF Binding Sites

130 While searching for interspecies differences in TF binding sites in *R1* and *R2*, we found evidence of a decreased  
 131 association of binding sites for *HOXB8*, *GATA-4*, *DLX3*, *CDX1*, *HOXA9*, *HOXB6*, *IRF4*, *ZNF384*, *TEF*, *MSX3*,  
 132 *MEOX1*, and *HSF2* with African green monkeys relative to other species in *R1*, as well as a decreased association  
 133 of binding sites for *ELF-1*, *GLS3*, *HOXB9*, *S MARCA2*, *OLF-1*, and *TIEG* with African green monkeys relative to  
 134 other species in *R2*.

Additionally, we found an increased association of binding sites for *GATA-3*, *MEIS1*, *PU.1/SPI-1*, and *EBF1* in *R1*, and an increased association of binding sites for *GCM1*, *GLIS2*, *KLF6*, *MAZ*, *MTBF*, *NM23*, *PAX6*, *RTT*, *SRF*, *STAT-1*, *STAT-3*, *ZIC2*, and *ZNF219* in *R2*, with African green monkeys relative to other species.

Some of the SNPs we have identified might be responsible for interspecies differences in *iNOS* regulation and expression, as well as susceptibility to diseases linked to *iNOS* expression levels, such as tuberculosis. While there is no clear biological link between many of the genes cited above and *iNOS* production, several of these genes (e.g. *GATA-3*, *GATA-4*, *KLF6*, *SRF*, *STAT-1*, *STAT-3*, *OLF-1* and *HIF-1*) have been linked to *iNOS* regulation in laboratory studies [33].

### 3.3. Candidate Genes for Interspecies Differences in *iNOS* Regulation

#### 3.3.1. STAT

Given the significant role of *STAT* proteins in inducing IFN-dependent expression of MHC II in macrophages [45], interspecies variation in *STAT* binding sites might contribute to variation in innate immune response across these species. *STAT-1*, specifically, is known to be an important transcription factor for regulating *iNOS* expression [46, 47]. Since several studies indicate a significant role for *STAT-1* in *iNOS* activation and expression [48], variation in *STAT-1* binding sites across these species might contribute to variable susceptibility to respiratory diseases, given the role of *iNOS* expression in immune response to respiratory infection [14]. Moreover, *STAT-1* plays a key role in inducing the expression of IFN-inducible 10kD protein (*IP-10*) and interferon regulatory factor 1 (*IRF-1*), which are also known to play key roles in innate immune response by macrophages [47]. Future *in vivo* investigations may clarify the influence of genetic changes in *STAT-1* binding sites in *iNOS* regulatory regions across non-human primates on susceptibility to respiratory diseases.

#### 3.3.2. KLF

Among these species of NHP, African green monkeys, but not members of the other species, exhibit a binding site for kruppel-like factor 6 (*KLF6*), a key transcription factor of *iNOS*. Several studies indicate that *KLF6*, which transactivates *iNOS* expression [49], plays a key role in immunity against viral and bacterial respiratory infections. A direct interaction between *KLF6* and *iNOS* has been observed in *in vitro* infection of human lung cells with influenza A virus [50]. *KLF6* also regulates apoptosis through the activation of *iNOS* expression and plays a critical role in *iNOS* expression during respiratory syncytial virus (RSV) [51] and influenza A virus infection [50].

### 3.4. Binding Site Differences in African Green Monkeys

The African green monkey has been widely used as an animal model for viral and bacterial respiratory diseases (e.g., RSV and *Mycobacterium tuberculosis* (MTB)) [52, 53, 54, 20] to which rhesus and cynomolgus macaques seem less susceptible. For example, African green monkeys exhibit more severe symptoms when infected with MTB [20]. Previous studies also indicate that replication of SARS virus occurs more rapidly in African green monkeys than in cynomolgus and rhesus macaques [21].

Because of the key role of *iNOS* in innate immune response against respiratory disease pathogens and the unique presence of binding sites for *STAT-1*, *KLF6*, and other *iNOS* REs/TFs in African Green monkeys relative to the other considered species, these RE/TF binding sites may play a role in the varying levels of susceptibility to viral and bacterial respiratory infections across these NHP species. However, variation in susceptibility to respiratory diseases is clearly multifactorial and regulated by many genes. Variation in binding sites for the above REs/TFs (i.e., *STAT-1* and *KLF6*), however, might partially explain variation in *iNOS* expression, and subsequently variation in innate immune response to respiratory infections.

More detailed laboratory studies are required to investigate: 1) if the RE/TF binding sites discovered through our *in silico* analysis are actually representative of biologically-relevant, *in vivo*, binding sites, 2) how these binding sites, if biologically relevant, modulate *iNOS* expression across species of NHP, and 3) how variation in *iNOS* expression is related to variation in disease susceptibility across species.

179 **4. Materials and Methods**

180 *4.1. Study Subjects, Sample Preparation, and DNA extraction*

181 DNA was extracted from lung tissue of 12 cynomolgus macaques (*Macaca fascicularis*) from the Tulane National  
 182 Primate Research Center that were previously used in a tuberculosis study [14], and from blood drawn from 5 Indian  
 183 and 5 Chinese rhesus macaques (*Macaca mulatta*) from the California National Primate Research Center, 5 pig-tailed  
 184 macaques (*Macaca nemestrina*) from Johns Hopkins University, and 6 African green monkeys (*Chlorocebus sabaeus*)  
 185 from the Wake Forest Primate Center using Qiagen QIAamp DNA Mini Kit (Qiagen Inc., Valencia, CA) following  
 186 the manufacturer's protocol.

187 *4.2. Candidate Promoter Regions of iNOS*

188 The first candidate regulatory region for *iNOS* that we label *R1*, for convenience, had been previously described  
 189 in the literature [see 55].

190 A second candidate regulatory region for *iNOS* that we label *R2* was mapped using the UCSC genome browser.  
 191 The *R2* region lays approximately 700-1000bp upstream of the first exon of *iNOS* and was identified as a regulatory  
 192 region candidate through visual inspection of the UCSC genome browser (see Figure 1) and literature review [see  
 193 33]. This region corresponds to the peak of *H3K4me3* in normal lung fibroblasts, and according to the UCSC genome  
 194 browser, contains several transcription factor binding sites (TFBS). The location of the orthologous sequence in NHP  
 195 species was located using the UCSC genome browser and the Basic Local Alignment Tool (BLAT).

196 *4.2.1. Primer Design and Sequencing of Two Possible Regulatory Regions of iNOS*

197 Two sets of primers were designed and tested to amplify the *R1* candidate regulatory region for *iNOS* [55]. Primer  
 198 set 1 was composed of forward and reverse primers 5'GGCAATGAGTGGACACTGGCA3' and 5'TGGCACAGAGATGCCCTCTGAGA3'  
 199 respectively, and primer set 2 was composed of forward and reverse primers 5'GGAAGGCAATGAGTGGACACTGGC3' and 5'GCTTTGGCAGAACATGGCAAGTAGGA3', respectively.

200 Two sets of primers were used to amplify the *R2* region: primer set 1 was composed of forward and reverse primers  
 201 5'CTACAGGTGAGTACACCCAGGAGCA3' and 5'GGCCTGTCCACCCTGGAGTGA3', respectively, and primer  
 202 set 2 was composed of forward and reverse primers 5'CCCAGGAGCAAGGAGAGGTGACA3' and 5'TGACTCACGCCCTCCAGTGAGA3', respectively.

203 The primers described above were tested using a gradient to optimize the annealing temperature for the polymerase  
 204 chain reaction (PCR). The PCR conditions used to amplify these two candidate regions for 60 PCR cycles were: an  
 205 initial hold at 94°C for 3 minutes, followed by denaturing at 94°C for 30 seconds, annealing at either 58.7°C for 20  
 206 seconds (for *R2*) or 57.4°C for 20 seconds (for *R1*), extension at 72°C for 45 seconds, and a final hold at 72°C for 5  
 207 minutes.

208 The amplified products were first quantified and checked by agarose and native acrylamide gel electrophoresis.  
 209 DNA sequencing was accomplished using the ABI BigDye Terminator sequencing chemistry and an ABI 3130XL  
 210 DNA sequencer (Applied Biosystems, Inc., Foster City, CA, USA) at the Molecular Anthropology Laboratory (MAL)  
 211 at UC Davis.

212 *4.3. Multiple Sequence Alignment*

213 Multiple sequence alignment was conducted with the command line version of Kalign2 [24, 25, 26] using several  
 214 hundred parameter permutations. The command line version of Mumsa [25] was used to select the best performing  
 215 alignments. The best performing alignments were then visually inspected and the procedure was repeated until a  
 216 sequence alignment with no noticeable pathologies was obtained.

217 The parameters used to obtain the final alignments for *R1* were: gap open penalty = 10, gap extension penalty =  
 218 1, and terminal gap penalty = 3. For *R2*, the final alignment parameters were: gap open penalty = 40, gap extension  
 219 penalty = 1, and terminal gap penalty = 6.

220 The Jalview program [30, 31] was used to visualize the sequence alignments and obtain consensus sequences.

221 *4.4. SNP Identification and Statistical Analysis*

222 The R Environment for Statistical Computing [27] was used for all SNP-based analysis of aligned sequences and  
 223 visualization of the results from the simulated regulatory element and transcription factor binding analysis.

#### 4.4.1. SNP Identification

SNPs were extracted from the aligned sequences using the *ape* [56] and *adegenet* [57] packages in R. Twenty-eight SNPs were identified in the trimmed version of *R1* (at locations 4, 7, 36, 68, 120, 135, 206, 288, 362, 394, 403, 406, 467, 478, 488, 492, 501, 509, 543, 550, 580, 592, 598, 599, 606, 612, 617, and 619 of the sequence). Twenty-eight SNPs were also identified in the trimmed version of *R2* (at locations 13, 19, 54, 81, 83, 101, 114, 123, 133, 155, 180, 222, 234, 238, 272, 287, 306, 323, 367, 386, 387, 405, 429, 485, 520, 591, 592, and 593 of the sequence). Tables 1 and 2 plot the sample frequencies of each SNP by species.

[Table 1 about here.]

[Table 2 about here.]

#### 4.4.2. Discriminant Analysis of Principal Components

The *adegenet* [57] package was utilized to conduct discriminant analysis of principal components [32] of the SNP data for each of the two sequence alignments. Further, *adegenet* was used to produce visualizations of these results both as scatter plots and group assignment probability plots.

#### 4.4.3. Simulated Transcription Factor Binding Analysis

We used the MatInspector software [28, 29] to search for RE and TF binding sites located inside *R1* and *R2*. This analysis used the most conservative default parameters settings in MatInspector and was conducted independently on each animal's aligned and trimmed FASTA formatted sequence.

MatInspector returns a large list of possible RE/TF binding sites for each DNA sequence. Because we are interested in understanding both conservation of and variation in RE/TF binding in *R1* and *R2* across species, we divided the full list of candidate RE/TF binding sites into two smaller list: 1) a list of RE/TF binding sites that did not include SNPs and were, therefore, conserved across species, and 2) a list of RE/TF binding sites that included at least one of the 28 SNPs and were, therefore, variable across animals/species.

#### 4.4.4. An $\ell_1$ -Regularized Categorical Bayesian Multiple Regression Model

For each region, *R1* and *R2*, we investigated how variation in RE/TF binding sites was associated with the classification of animals into the  $K = 3$  major clusters observed in our data: African Green monkeys, Chinese and Indian rhesus macaques, and cynomolgus and pig-tailed macaques. We used a categorical regression model to predict the classification,  $Y_{[n]} \in \{1 \dots K\}$ , of animal  $n$  using data on the RE/TF binding profile of that animal. Each RE/TF binding profile,  $X_{[1\dots(P+1),n]}$ , is a vector beginning with an intercept value of 1, and continuing with  $P = 105$  (in *R1*, or  $P = 95$  in *R2*) binary data points that indicate the presence or absence of the  $p_{th}$  RE/TF binding site in animal  $n$ . Because we had many more predictors than animals in our sample, we used a full Bayesian regression model with Laplace (also known as double exponential) priors on the regression coefficients. This model formulation imposes the Bayesian corollary of Lasso, or  $\ell_1$ -regularized regression, which penalizes the number of non-zero parameter values, reducing effective model complexity [58]. Accordingly, each outcome is modeled as:

$$Y_{[n]} \sim \text{Categorical}(\phi_{[n]}) \quad (1)$$

where:

$$\phi_{[n]} = \text{Softmax}(\beta * X_{[1\dots(P+1),n]}) \quad (2)$$

and  $\beta$  is a  $K$  by  $P + 1$  matrix of parameters representing the intercepts as well as the associations of each of the  $P$  predictors with each of the  $K$  outcome categories. The Softmax function is defined for a  $K$ -vector  $\theta \in \mathbb{R}^K$  as:

$$\text{Softmax}(\theta) = \left( \frac{e^{\theta_{[1]}}}{\sum_{k=1}^K e^{\theta_{[k]}}}, \dots, \frac{e^{\theta_{[K]}}}{\sum_{k=1}^K e^{\theta_{[k]}}} \right) \quad (3)$$

and yields a vector in the unit  $K$ -simplex which is appropriate to use as the parameter vector for a categorical distribution. The Softmax function is invariant under adding a constant to each component of its input [59], but strongly regularizing priors identify the model.

We declare weakly regularizing Gaussian priors on the intercept parameter for each category where  $k \in \{1 \dots K\}$ :

$$\beta_{[k,1]} \sim \text{Normal}(0, 10) \quad (4)$$

and strongly regularizing Laplace priors on the remaining (slope) parameters of the  $\beta$  matrix. Thus, for  $k \in \{1 \dots K\}$  and  $p \in \{2 \dots (P + 1)\}$ , we model:

$$\beta_{[k,p]} \sim \text{Double Exponential}(0, 0.1) \quad (5)$$

We use Hamiltonian Markov Chain Monte Carlo simulation [60] to fit this model. Our Markov chains are coded in templated C++ using the R implementation of the Stan 2.2.0 C++ library [61]. We then use the *vcf* package [62] in R to plot the posterior mean estimate of each  $\beta_{[1\dots K,p]}$  parameter vector in barycentric coordinates. Points located in the center of the barycentric space indicate variables that are not informative about inter-group differences in RE/TF binding sites, while points located near the edges of the barycentric space are indicative of inter-group differences in RE/TF binding sites.

## 5. Supplementary Material

Supplementary Tables 1, 2, and 3, as well as FASTA formatted sequences for each animal are available online.

## 6. Acknowledgments

The authors gratefully acknowledge the help of Frederic Chedin (Department of Molecular and Cellular Biology UC Davis Genome Center), Satya Dandekar (Medical Microbiology and Immunology School of Medicine, UC Davis, and Linda Lowenstein (Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine UC Davis).

This work was supported by National Institute of Health grants R24RR05090, R24RR025871, TR000002, and HL105573.

## References

- [1] J. Chan, Y. Xing, R. Magliozzo, B. Bloom, Killing of virulent mycobacterium tuberculosis by reactive nitrogen intermediates produced by activated murine macrophages., *The Journal of experimental medicine* 175 (4) (1992) 1111–1122.
- [2] J. MacMicking, Q.-w. Xie, C. Nathan, Nitric oxide and macrophage function, *Annual review of immunology* 15 (1) (1997) 323–350.
- [3] K. Nanashima, T. Mawatari, N. Tahara, N. Higuchi, A. Nakaura, T. Inamine, S. Kondo, K. Yanagihara, K. Fukushima, N. Suyama, et al., Genetic variants in antioxidant pathway: risk factors for hepatotoxicity in tuberculosis patients, *Tuberculosis* 92 (3) (2012) 253–259.
- [4] J. M. Park, M.-K. Baeg, C.-H. Lim, Y. K. Cho, M.-G. Choi, Nitric oxide synthase gene polymorphisms in functional dyspepsia, *Digestive diseases and sciences* 59 (1) (2014) 72–77.
- [5] Y.-P. Lim, C.-Y. Peng, W.-L. Liao, D.-Z. Hung, N. Tien, C.-Y. Chen, S.-Y. Chang, C.-Y. Chang, F.-J. Tsai, L. Wan, Genetic variation in nos2a is associated with a sustained virological response to peginterferon plus ribavirin therapy for chronic hepatitis c in taiwanese chinese, *Journal of medical virology* 85 (7) (2013) 1206–1214.
- [6] S. AlFadhli, E. M. Mohammed, A. Al Shubaili, Association analysis of nitric oxide synthases: Nos1, nos2a and nos3 genes, with multiple sclerosis, *Annals of human biology* 40 (4) (2013) 368–375.
- [7] A. Fabisiewicz, K. Pacholewicz, E. Paszkiewicz-Kozik, J. Walewski, J. A. Siedlecki, Polymorphisms of dna repair and oxidative stress genes in b-cell lymphoma patients, *Biomedical Reports* 1 (1) (2013) 151–155.
- [8] J. A. Karasneh, A. M. Darwazeh, A. F. Hassan, M. Thornhill, Association between recurrent aphthous stomatitis and inheritance of a single-nucleotide polymorphism of the nos2 gene encoding inducible nitric oxide synthase, *Journal of Oral Pathology & Medicine* 40 (9) (2011) 715–720.
- [9] Z. Wang, K. Feng, M. Yue, X. Lu, Q. Zheng, H. Zhang, Y. Zhai, P. Li, L. Yu, M. Cai, et al., A non-synonymous snp in the nos2 associated with septic shock in patients with sepsis in chinese populations, *Human genetics* 132 (3) (2013) 337–346.
- [10] A. Rafiei, V. Hosseini, G. Janbabai, B. Fazli, A. Ajami, Z. Hosseini-Khah, J. Gilbreath, D. S. Merrell, Inducible nitric oxide synthetase genotype and helicobacter pylori infection affect gastric cancer risk, *World journal of gastroenterology: WJG* 18 (35) (2012) 4917.
- [11] Y. Zhang, C. Li, K. Li, L. Liu, Z. Jian, T. Gao, Analysis of inducible nitric oxide synthase gene polymorphisms in vitiligo in han chinese people, *PloS one* 6 (12) (2011) e27077.

- [12] T. Planche, D. C. Macallan, T. Sobande, S. Borrmann, J. F. Kun, S. Krishna, P. G. Kremsner, Nitric oxide generation in children with malaria and the nos2g-954c promoter polymorphism, American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 299 (5) (2010) R1248–R1253.
- [13] M. C. Levesque, M. R. Hobbs, C. W. O'Loughlin, J. A. Chancellor, Y. Chen, A. N. Tkachuk, J. Booth, K. B. Patch, S. Allgood, A. R. Pole, et al., Malaria severity and human nitric oxide synthase type 2 (nos2) promoter haplotypes, Human genetics 127 (2) (2010) 163–182.
- [14] M. Roodgar, A. Lackner, D. Kaushal, S. Sankaran, S. Dandekar, J. Satkoski Trask, C. Drake, D. G. Smith, Expression levels of 10 candidate genes in lung tissue of vaccinated and tb-infected cynomolgus macaques, Journal of Medical Primatology 42 (3) (2013) 161–164.
- [15] T. J. Gross, K. Kremens, L. S. Powers, B. Brink, T. Knutson, F. E. Domann, R. A. Philibert, M. M. Milhem, M. M. Monick, Epigenetic silencing of the human nos2 gene: Rethinking the role of nitric oxide in human macrophage inflammatory responses, The Journal of Immunology (2014) 1301758.
- [16] M. C. de Andrés, K. Imagawa, K. Hashimoto, A. Gonzalez, H. I. Roach, M. B. Goldring, R. O. Oreffo, Loss of methylation in cpg sites in the nf- $\kappa$ b enhancer elements of inducible nitric oxide synthase is responsible for gene induction in human articular chondrocytes, Arthritis & Rheumatism 65 (3) (2013) 732–742.
- [17] T. Angrisano, F. Lembo, S. Peluso, S. Keller, L. Chiariotti, R. Pero, Helicobacter pylori regulates inos promoter by histone modifications in human gastric epithelial cells, Medical microbiology and immunology 201 (3) (2012) 249–257.
- [18] S. Jia, J. Ni, S. Chen, Y. Jiang, W. Dong, Y. Gao, Association of the pentanucleotide repeat polymorphism in nos2 promoter region with susceptibility to migraine in a chinese population, DNA and cell biology 30 (2) (2011) 117–122.
- [19] M. R. Hobbs, V. Udhayakumar, M. C. Levesque, J. Booth, J. M. Roberts, A. N. Tkachuk, A. Pole, H. Coon, S. Kariuki, B. L. Nahnen, et al., A new i<sub>2</sub>; nos2i<sub>2</sub>/i<sub>2</sub> promoter polymorphism associated with increased nitric oxide production and protection from severe malaria in tanzanian and kenyian children, The Lancet 360 (9344) (2002) 1468–1475.
- [20] K. P. Lyashchenko, R. Greenwald, J. Esfandiari, D. Greenwald, C. A. Nacy, S. Gibson, P. J. Didier, M. Washington, P. Szczerba, S. Motzel, et al., Primath stat-pak assay, a novel, rapid lateral-flow test for tuberculosis in nonhuman primates, Clinical and Vaccine Immunology 14 (9) (2007) 1158–1164.
- [21] J. McAuliffe, L. Vogel, A. Roberts, G. Fahle, S. Fischer, W.-J. Shieh, E. Butler, S. Zaki, M. St Claire, B. Murphy, et al., Replication of sars coronavirus administered into the respiratory tract of african green, rhesus and cynomolgus monkeys, Virology 330 (1) (2004) 8–15.
- [22] S. Wienerroither, I. Rauch, F. Rosebrock, A. M. Jamieson, J. Bradner, M. Muhar, J. Zuber, M. Müller, T. Decker, Regulation of no synthesis, local inflammation, and innate immunity to pathogens by bet family proteins, Molecular and cellular biology 34 (3) (2014) 415–427.
- [23] N. Obermajer, J. L. Wong, R. P. Edwards, K. Chen, M. Scott, S. Khader, J. K. Kolls, K. Odunsi, T. R. Billiar, P. Kalinski, Induction and stability of human th17 cells require endogenous nos2 and cgmp-dependent no signaling, The Journal of experimental medicine 210 (7) (2013) 1433–1445.
- [24] T. Lassmann, O. Frings, E. L. Sonnhammer, Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features, Nucleic acids research 37 (3) (2009) 858–865.
- [25] T. Lassmann, E. L. Sonnhammer, Kalign, kalignvu and mumsa: web servers for multiple sequence alignment, Nucleic acids research 34 (suppl 2) (2006) W596–W599.
- [26] T. Lassmann, E. L. Sonnhammer, Kalign—an accurate and fast multiple sequence alignment algorithm, BMC bioinformatics 6 (1) (2005) 298.
- [27] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2014). URL <http://www.R-project.org>
- [28] K. Quandt, K. Frech, H. Karas, E. Wingender, T. Werner, MatInd and matInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data, Nucleic acids research 23 (23) (1995) 4878–4884.
- [29] K. Cartharius, K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein, T. Werner, MatInspector and beyond: promoter analysis based on transcription factor binding sites, Bioinformatics 21 (13) (2005) 2933–2942.
- [30] M. Clamp, J. Cuff, S. M. Searle, G. J. Barton, The jalview java alignment editor, Bioinformatics 20 (3) (2004) 426–427.
- [31] A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, Jalview version 2a multiple sequence alignment editor and analysis workbench, Bioinformatics 25 (9) (2009) 1189–1191.
- [32] T. Jombart, S. Devillard, F. Balloux, Discriminant analysis of principal components: a new method for the analysis of genetically structured populations, BMC genetics 11 (1) (2010) 94.
- [33] A. Pautz, J. Art, S. Hahn, S. Nowag, C. Voss, H. Kleinert, Regulation of the expression of inducible nitric oxide synthase, Nitric Oxide 23 (2) (2010) 75–93.
- [34] X. Liao, N. Sharma, F. Kapadia, G. Zhou, Y. Lu, H. Hong, K. Paruchuri, G. H. Mahabeleshwar, E. Dalmas, N. Venteclef, et al., Krüppel-like factor 4 regulates macrophage polarization, The Journal of clinical investigation 121 (7) (2011) 2736.
- [35] K. Hayasaka, K. Fujii, S. Horai, Molecular phylogeny of macaques: Implications of nucleotide sequences from an 896-base pair region of mitochondrial dna., Molecular Biology and Evolution 13 (7) (1996) 1044–1053.
- [36] A. J. Tosi, J. C. Morales, D. J. Melnick, Comparison of y chromosome and mtDNA phylogenies leads to unique inferences of macaque evolutionary history, Molecular phylogenetics and evolution 17 (2) (2000) 133–144.
- [37] H. Kleinert, A. Pautz, K. Linker, P. M. Schwarz, Regulation of the expression of inducible nitric oxide synthase, European journal of pharmacology 500 (1) (2004) 255–266.
- [38] D. G. Uffort, E. A. Grimm, J. A. Ellerhorst, Nf- $\kappa$ b mediates mitogen-activated protein kinase pathway-dependent inos expression in human melanoma, Journal of Investigative Dermatology 129 (1) (2009) 148–154.
- [39] E. D. Chan, K. R. Morris, J. T. Belisle, P. Hill, L. K. Remigio, P. J. Brennan, D. W. Riches, Induction of inducible nitric oxide synthase-no- by lipoarabinomannan of mycobacterium tuberculosis is mediated by mek1-erk, mkk7-jnk, and nf- $\kappa$ b signaling pathways, Infection and immunity 69 (4).
- [40] M. Jaramillo, D. C. Gowda, D. Radzioch, M. Olivier, Hemozoin increases ifn- $\gamma$ -inducible macrophage nitric oxide generation through extracellular signal-regulated kinase-and nf- $\kappa$ b-dependent pathways, The Journal of Immunology 171 (8) (2003) 4243–4253.
- [41] L. Oussaief, V. Ramírez, A. Hippocrate, H. Arbach, C. Cochet, A. Proust, M. Raphaël, R. Khelifa, I. Joab, Nf- $\kappa$ b-mediated modulation of inducible nitric oxide synthase activity controls induction of the epstein-barr virus productive cycle by transforming growth factor beta 1,

- 372 Journal of virology 85 (13) (2011) 6502–6512.
- 373 [42] Z. T. Kelleher, A. Matsumoto, J. S. Stamler, H. E. Marshall, Nos2 regulation of nf- $\kappa$ b by s-nitrosylation of p65, Journal of biological chemistry 282 (42) (2007) 30667–30672.
- 374 [43] X. Feng, Z. Guo, M. Nourbakhsh, H. Hauser, R. Ganster, L. Shao, D. A. Geller, Identification of a negative response element in the human inducible nitric-oxide synthase (inos) promoter: the role of nf- $\kappa$ b-repressing factor (nrf) in basal repression of the inos gene, Proceedings of the National Academy of Sciences 99 (22) (2002) 14212–14217.
- 375 [44] J. M. Kim, J. S. Kim, H. C. Jung, Y.-K. Oh, H.-Y. Chung, C.-H. Lee, I. S. Song, Helicobacter pylori infection activates nf- $\kappa$ b signaling pathway to induce inos and protect human gastric epithelial cells from apoptosis, American Journal of Physiology-Gastrointestinal and Liver Physiology 285 (6) (2003) G1171–G1180.
- 376 [45] W. Zhao, E. N. Cha, C. Lee, C. Y. Park, C. Schindler, Stat2-dependent regulation of mhc class ii expression, The Journal of Immunology 179 (1) (2007) 463–471.
- 377 [46] R. W. Ganster, B. S. Taylor, L. Shao, D. A. Geller, Complex regulation of human inducible nitric oxide synthase gene transcription by stat 1 and nf- $\kappa$ b, Proceedings of the National Academy of Sciences 98 (15) (2001) 8638–8643.
- 378 [47] Y. Ohmori, T. A. Hamilton, Requirement for stat1 in lps-induced gene expression in macrophages, Journal of Leukocyte Biology 69 (4) (2001) 598–604.
- 379 [48] T. Samardzic, V. Jankovic, S. Stosic-Grujicic, V. Trajkovic, Stat1 is required for inos activation, but not il-6 production in murine fibroblasts, Cytokine 13 (3) (2001) 179–182.
- 380 [49] V. G. Warke, M. P. Nambiar, S. Krishnan, K. Tenbrock, D. A. Geller, N. P. Koritschoner, J. L. Atkins, D. L. Farber, G. C. Tsokos, Transcriptional activation of the human inducible nitric-oxide synthase promoter by krüppel-like factor 6, Journal of Biological Chemistry 278 (17) (2003) 14812–14819.
- 381 [50] V. Mgbemena, J. A. Segovia, T.-H. Chang, S.-Y. Tsai, G. T. Cole, C.-Y. Hung, S. Bose, Transactivation of inducible nitric oxide synthase gene by krüppel-like factor 6 regulates apoptosis during influenza a virus infection, The Journal of Immunology 189 (2) (2012) 606–615.
- 382 [51] V. Mgbemena, J. Segovia, T.-H. Chang, S. Bose, Klf6 and inos regulates apoptosis during respiratory syncytial virus infection, Cellular immunology 283 (1) (2013) 1–7.
- 383 [52] A. Bukreyev, E. W. Lamirande, U. J. Buchholz, L. N. Vogel, W. R. Elkins, M. S. Claire, B. R. Murphy, K. Subbarao, P. L. Collins, Mucosal immunisation of african green monkeys (*l. tarsi* cercopithecus aethiops)/*l. tarsi* with an attenuated parainfluenza virus expressing the sars coronavirus spike protein for the prevention of sars, The Lancet 363 (9427) (2004) 2122–2127.
- 384 [53] R. S. Tang, M. MacPhail, J. H. Schickli, J. Kaur, C. L. Robinson, H. A. Lawlor, J. M. Guzzetta, R. R. Spaete, A. A. Haller, Parainfluenza virus type 3 expressing the native or soluble fusion (f) protein of respiratory syncytial virus (rsv) confers protection from rsv infection in african green monkeys, Journal of virology 78 (20) (2004) 11198–11207.
- 385 [54] H. Jin, X. Cheng, V. L. Traina-Dorge, H. J. Park, H. Zhou, K. Soike, G. Kemble, Evaluation of recombinant respiratory syncytial virus gene deletion mutants in african green monkeys for their potential as live attenuated vaccine candidates, Vaccine 21 (25) (2003) 3647–3652.
- 386 [55] Y. Nunokawa, N. Ishida, S. Tanaka, Promoter analysis of human inducible nitric oxide synthase gene associated with cardiovascular homeostasis, Biochemical and Biophysical Research Communications 200 (2) (1994) 802 – 807. doi:<http://dx.doi.org/10.1006/bbrc.1994.1522>.  
URL <http://www.sciencedirect.com/science/article/pii/S0006291X84715221>
- 387 [56] E. Paradis, J. Claude, K. Strimmer, APE: analyses of phylogenetics and evolution in R language, Bioinformatics 20 (2004) 289–290.
- 388 [57] T. Jombart, adegenet: a r package for the multivariate analysis of genetic markers, Bioinformatics 24 (11) (2008) 1403–1405.
- 389 [58] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) (1996) 267–288.
- 390 [59] Stan Development Team, Stan Modeling Language Users Guide and Reference Manual, Version 2.2 (2014).  
URL <http://mc-stan.org/>
- 391 [60] M. D. Hoffman, A. Gelman.
- 392 [61] Stan Development Team, Stan: A c++ library for probability and sampling, version 2.0 (2013).  
URL <http://mc-stan.org/>
- 393 [62] D. Meyer, A. Zeileis, K. Hornik, vcd: Visualizing categorical data, R package version (2006) 1–0.

**List of Figures**

	417
1 Location of the <i>iNOS</i> coding region and the <i>R2</i> candidate promoter/regulatory region on the human reference frame. . . . .	418 12
2 Results of multiple sequence alignment in regions <i>R1</i> (Frame 2a) and <i>R2</i> (Frame 2b). . . . .	419 13
3 Discriminant analysis of principal components in regions <i>R1</i> (Frame 3a) and <i>R2</i> (Frame 3b). The first principal component (horizontal axis) separates African green monkeys from the other species in both <i>R1</i> and <i>R2</i> . The second principal component (vertical axis) separates Rhesus macaques from cynomolgus and pig-tailed macaques. . . . .	420 421 422 423 14
4 Group assignment plots based on DAPC analysis of regions <i>R1</i> (Frame 4a) and <i>R2</i> (Frame 4b). The row labels are animal IDs (the numeric values are African green monkeys, the <i>Cay</i> prefix indicates Indian rhesus macaques, the <i>Jhop</i> prefix indicates pig-tailed macaques, the <i>FG</i> and <i>FB</i> prefixes are cynomolgus macaques, and the <i>Sch</i> prefix indicates Chinese rhesus macaques). The blue crosses indicate the true group assignment, and the cell color indicates increasing group assignment probability (based on SNP data) as the scale shifts from yellow to red. We note three distinct genetic clusters in <i>R1</i> (African green monkey, rhesus macaque, and cynomolgus/pig-tailed macaques), and four distinct genetic clusters in <i>R2</i> (African green monkey, Indian rhesus macaque, Chinese rhesus macaque, and cynomolgus/pig-tailed macaques). . . . .	424 425 426 427 428 429 430 431 432 15 433
5 The associations between RE and TF binding sites and species classification in regions <i>R1</i> (Frame 5a) and <i>R2</i> (Frame 5b). Each colored letter represents an RE/TF binding site (see Supplementary Materials Table 3 for the codebook linking each RE/TF binding site to each data point in the plot) plotted in barycentric space using a ternary plot to represent the relative strength of association across categories. Data points in the center cluster of the plotting space are not indicative of interspecies differences in RE/TF binding sites; data points located near the boundary of the triangle are indicative of interspecies differences in RE/TF binding. Only a small fraction of the RE/TF binding sites identified in our analysis are indicative of interspecies differences. . . . .	434 435 436 437 438 439 440 16 441

Figure 1: Location of the *iNOS* coding region and the *R2* candidate promoter/regulatory region on the human reference frame.

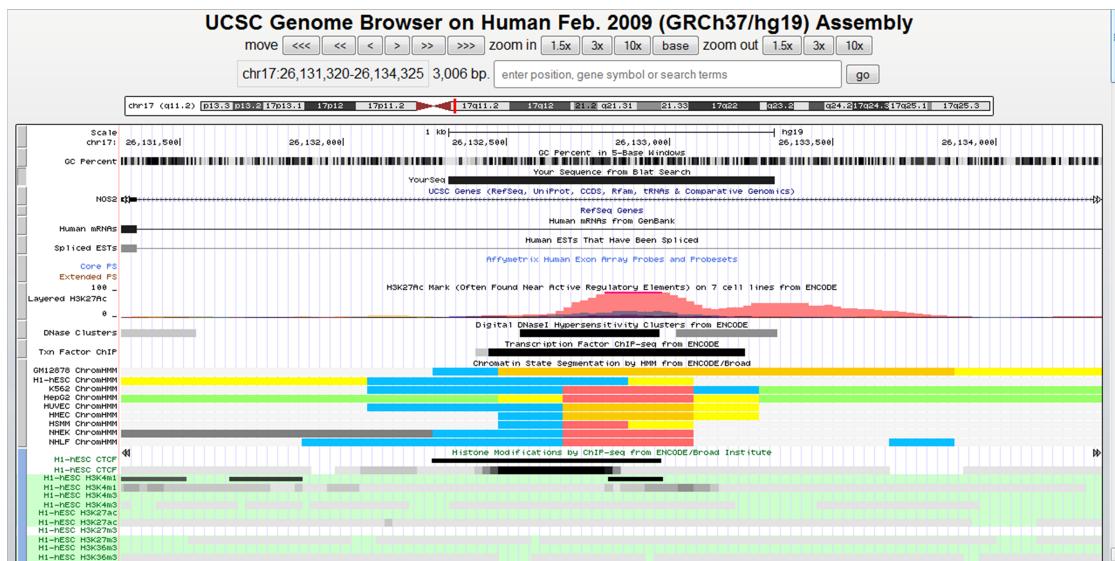
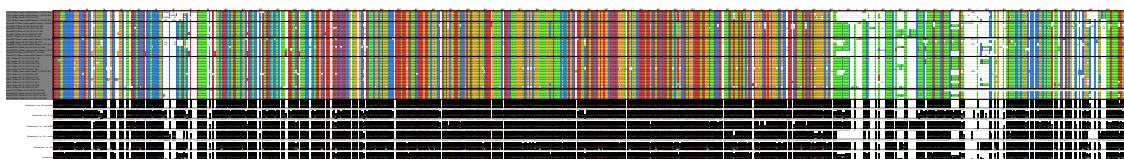
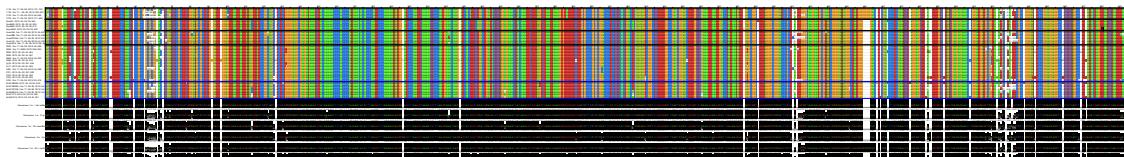


Figure 2: Results of multiple sequence alignment in regions  $R1$  (Frame 2a) and  $R2$  (Frame 2b).



(a) Alignment and Consensus sequences of  $R1$



(b) Alignment and Consensus sequences of  $R2$

Figure 3: Discriminant analysis of principal components in regions  $R1$  (Frame 3a) and  $R2$  (Frame 3b). The first principal component (horizontal axis) separates African green monkeys from the other species in both  $R1$  and  $R2$ . The second principal component (vertical axis) separates Rhesus macaques from cynomolgus and pig-tailed macaques.

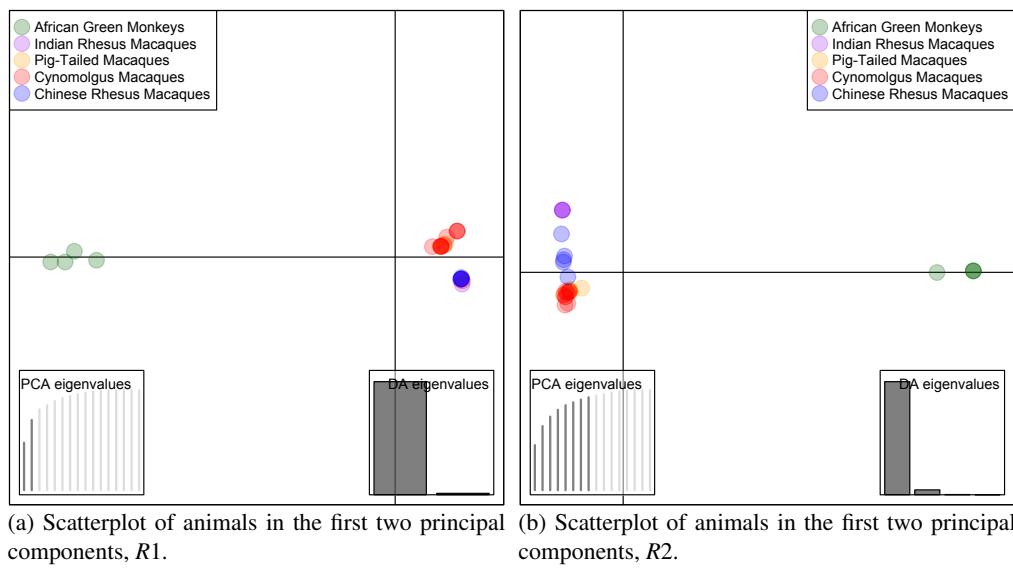
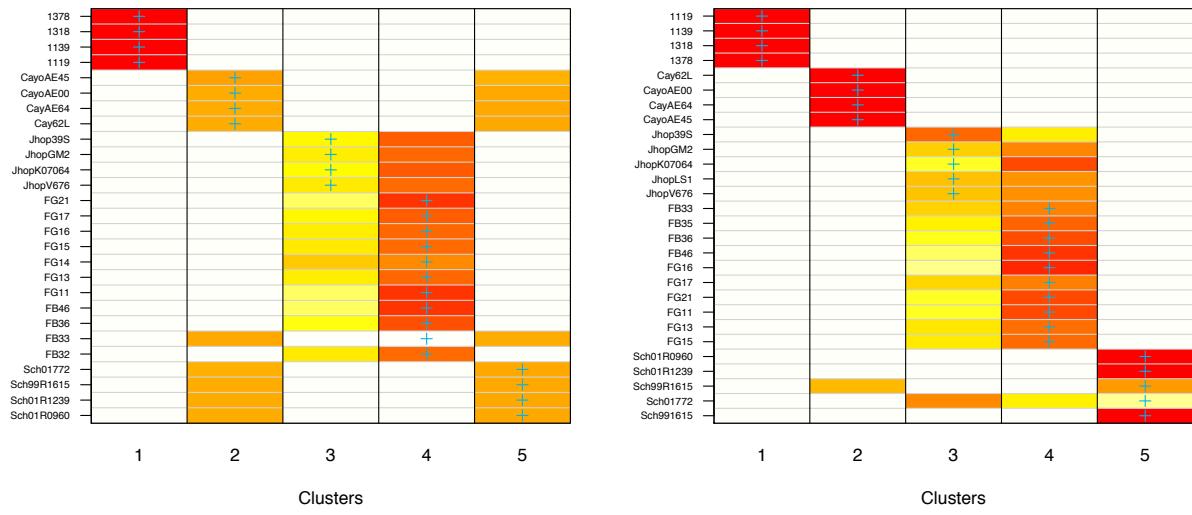
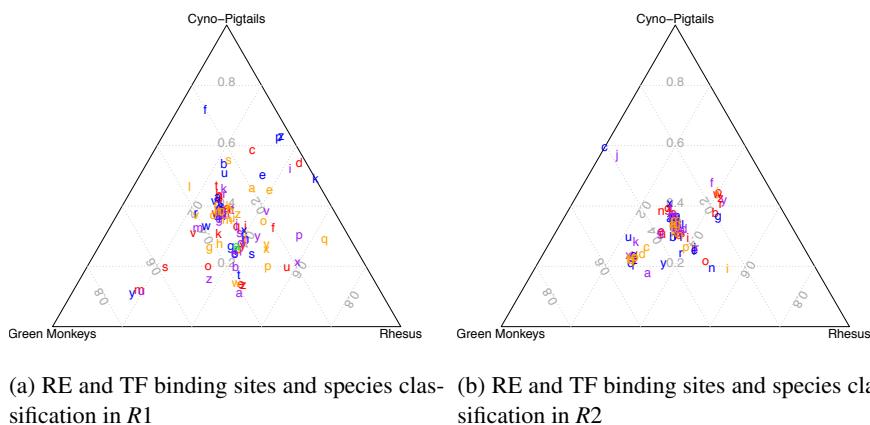


Figure 4: Group assignment plots based on DAPC analysis of regions  $R1$  (Frame 4a) and  $R2$  (Frame 4b). The row labels are animal IDs (the numeric values are African green monkeys, the *Cay* prefix indicates Indian rhesus macaques, the *Jhop* prefix indicates pig-tailed macaques, the *FG* and *FB* prefixes are cynomolgus macaques, and the *Sch* prefix indicates Chinese rhesus macaques). The blue crosses indicate the true group assignment, and the cell color indicates increasing group assignment probability (based on SNP data) as the scale shifts from yellow to red. We note three distinct genetic clusters in  $R1$  (African green monkey, rhesus macaque, and cynomolgus/pig-tailed macaques), and four distinct genetic clusters in  $R2$  (African green monkey, Indian rhesus macaque, Chinese rhesus macaque, and cynomolgus/pig-tailed macaques).



(a) Group assignment probability plot of animals using region  $R1$ . (b) Group assignment probability plot of animals using region  $R2$ .

Figure 5: The associations between RE and TF binding sites and species classification in regions  $R1$  (Frame 5a) and  $R2$  (Frame 5b). Each colored letter represents an RE/TF binding site (see Supplementary Materials Table 3 for the codebook linking each RE/TF binding site to each data point in the plot) plotted in barycentric space using a tertiary plot to represent the relative strength of association across categories. Data points in the center cluster of the plotting space are not indicative of interspecies differences in RE/TF binding sites; data points located near the boundary of the triangle are indicative of interspecies differences in RE/TF binding. Only a small fraction of the RE/TF binding sites identified in our analysis are indicative of interspecies differences.



**List of Tables**

1	Sample SNP frequencies by species in region <i>R</i> 1. . . . .	18	443
2	Sample SNP frequencies by species in region <i>R</i> 2. . . . .	19	444

Table 1: Sample SNP frequencies by species in region R1.

SNP Location	African Green Monkey	Indian Rhesus Macaque	Pig-tailed Macaque	Cynomolgus Macaque	Chinese Rhesus Macaque
4	0	0	0	0.083	0
7	0	1	1	1	1
36	0	1	1	1	1
68	0.333	1	1	1	1
120	0	0	0.143	0	0
135	NA	NA	0.667	NA	NA
206	0	1	1	1	1
288	0	1	0	0.083	1
362	0	1	0	0.083	1
394	0.5	1	1	1	1
403	0	1	1	1	1
406	0	1	1	1	1
467	NA	0	1	NA	NA
478	0	0	0.143	0	0
488	1	1	0.857	1	1
492	NA	1	0	1	1
501	1	1	0.715	1	1
509	1	0	0	0	0
543	0.75	1	1	1	1
550	1	0.75	1	1	1
580	1	1	1	0.917	1
592	1	0.667	1	1	1
598	1	1	0.857	1	1
599	1	1	0.857	1	1
606	0.75	1	1	0.667	1
612	1	0.833	1	1	1
617	0.75	1	1	1	1
619	1	1	0.857	1	1

Table 2: Sample SNP frequencies by species in region R2.

SNP Location	African Green Monkey	Indian Rhesus Macaque	Pig-tailed Macaque	Cynomolgus Macaque	Chinese Rhesus Macaque
13	0	0	0	0	0.167
19	0	0	0	0.077	0
54	0	1	1	1	1
81	0	0	0.25	0	0
83	0	0	0	0.077	0
101	0	1	1	1	1
114	0	1	0	0	0.5
123	0	0	0.25	0	0
133	0	0	0	0.25	0
155	0	0	0.2	0	0
180	0	1	1	1	1
222	0	0	0.333	0	0
234	0	1	1	1	1
238	0	1	0	0	0
272	0	0	0	0	0.333
287	0	1	1	1	1
306	0	0	0	0.154	0
323	0	1	0	0	0.4
367	0	0	0	0.077	0
386	0	0	0	0.5	0
387	0	1	1	1	1
405	0	1	1	1	1
429	0	1	1	1	1
485	0	0	0	0.154	0.2
520	0	0	0	0.667	0
591	0	0	0.2	0.077	0
592	0	0	0	0.077	0
593	0	0	0.2	0	0