

# SCIENTIFIC REPORTS



OPEN

## Non-invasive genotyping with a massively parallel sequencing panel for the detection of SNPs in HPA-axis genes

D. R. Gutleb<sup>1,2,3</sup> , J. Ostner<sup>1,2,3</sup>, O. Schülke<sup>1,2,3</sup>, W. Wajjwalku<sup>4</sup>, M. Sukmak<sup>4</sup>, C. Roos<sup>5,6</sup>  & A. Noll<sup>6</sup>

We designed a genotyping panel for the investigation of the genetic underpinnings of inter-individual differences in aggression and the physiological stress response. The panel builds on single nucleotide polymorphisms (SNPs) in genes involved in the three subsystems of the hypothalamic-pituitary-adrenal (HPA)-axis: the catecholamine, serotonin and corticoid metabolism. To promote the pipeline for use with wild animal populations, we used non-invasively collected faecal samples from a wild population of Assamese macaques (*Macaca assamensis*). We targeted loci of 46 previously reported SNPs in 21 candidate genes coding for elements of the HPA-axis and amplified and sequenced them using next-generation Illumina sequencing technology. We compared multiple bioinformatics pipelines for variant calling and variant effect prediction. Based on this strategy and the application of different quality thresholds, we identified up to 159 SNPs with different types of predicted functional effects among our natural study population. This study provides a massively parallel sequencing panel that will facilitate integrating large-scale SNP data into behavioural and physiological studies. Such a multi-faceted approach will promote understanding of flexibility and constraints of animal behaviour and hormone physiology.

Recent developments in molecular techniques enable researchers to include large-scale investigations of genetic impacts on behavioural or endocrine variables at reasonable costs<sup>1,2</sup>. In studies on humans, the investigation of genotypic influences on phenotypic characteristics revealed that inter-individual variation can be strongly affected by genotype<sup>3-5</sup>. For example, genotypic variation is responsible for approximately 50% of inter-individual variation in physiological stress levels and aggression<sup>6,7</sup>. In animal studies, however, the consideration of underlying genotype in behavioural and physiological studies is relatively understudied. Especially regarding studies on wild, non-model populations, several authors have called for a more frequent consideration of genetic impacts on animal behaviour<sup>1,8,9</sup>. In this study, we provide a multi-locus genotyping pipeline, based on non-invasively sampled material from a population of wild primates, facilitating future research on social and ecological factors driving variation in stress and aggression.

Modulation of the hypothalamic-pituitary-adrenal (HPA)-axis activity is an effective mechanism mediating environmental effects on the organism including its behavioural tendencies. The HPA-axis is a central physiological pathway activated in response to stress and is conserved across vertebrates<sup>10,11</sup>. In behavioural ecological studies, aggressive behaviour is often linked to HPA-axis activity via quantification of cortisol, the end product of this pathway (fish<sup>12</sup>, birds<sup>13</sup>, rodents<sup>14</sup>, ruminants<sup>15</sup>, cats and dogs<sup>16,17</sup>, primates<sup>18,19</sup>). Links to behaviour have been established in rats where the stimulation of brain areas responsible for aggression causes cortisol release,

<sup>1</sup>Department of Behavioral Ecology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology, University of Goettingen, Göttingen, Germany. <sup>2</sup>Research Group Social Evolution in Primates, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany. <sup>3</sup>Leibniz Science Campus Primate Cognition, Göttingen, Germany. <sup>4</sup>Department of Farm Resources and Production Medicine, Faculty of Veterinary Medicine, Kasetsart University, Nakhon Pathom, Thailand. <sup>5</sup>Gene Bank of Primates, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany. <sup>6</sup>Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany. Correspondence and requests for materials should be addressed to D.R.G. (email: [daria@gutleb.at](mailto:daria@gutleb.at))

and similar processes are proposed for other vertebrates, including humans<sup>20–22</sup>. Behavioural ecological studies on aggression often assess how the expression of aggressive behaviour is related to social dominance or affected by characteristics of the competitive situation without conceptually integrating inter-individual variation due to genetic variation.

Three main metabolic circuits contribute to the HPA-axis: the serotonin, the catecholamine and the corticoid. Genes associated with these metabolic circuits have been repeatedly targeted in human clinical stress<sup>23,24</sup> and aggression<sup>4,25,26</sup> research. Functional polymorphisms in the genes coding for the three subsystems may lead to dysregulations in the HPA-axis pathway and a change in how the organism reacts to external stressors. The serotonin pathway involves the serotonin transporter (SLC6A4), receptor (HTR), tryptophan hydroxylase (TPH) and monoamine oxidase<sup>27</sup>. The neuropeptide Y (NPY) is a neurochemical that plays a protective role in stress resilience<sup>24,28</sup>. The catecholamine circuit (e.g. dopamine: DRD, SLC6A3, catechol-O-methyl transferase: COMT) causes general physiological changes that prepare the body for physical activity<sup>29</sup>. Main components of the corticoid pathway are the corticotropin-releasing hormone (CRH), CRH receptors (CRHR), the glucocorticoid receptor (NR3C1), CRH binding protein (CRHBP), corticosteroid binding globulin (SERPINA6) and the corticotropin receptor (MC2R)<sup>30</sup>.

For human diagnostics in the field of behavioural genetics, studies acquire large datasets via high-throughput methods such as massively parallel sequencing (i.e. next-generation sequencing)<sup>2</sup>. Behavioural studies on natural animal populations, however, commonly target one or a few gene loci associated with aggression and HPA-axis activity<sup>31–33</sup>, whereas high-throughput multi-locus approaches are rather rare but see<sup>1,34</sup>. Due to the high number of genes involved in physiological pathways such as the HPA-axis, the simultaneous assessment of multiple loci known to affect certain traits promises a much more comprehensive understanding of the investigated physiological and behavioural parameters<sup>35,36</sup>. The introduction of massively parallel sequencing technologies makes a multi-locus approach also feasible in studies on non-model species. The generated data provide high coverage of amplicons or genomes and a large and still growing body of different bioinformatics applications helps to investigate multiple loci in a fast and parallel way (e.g. Genome Analysis Toolkit – GATK<sup>37</sup>, SAMtools<sup>38</sup> and UCSC genome browser<sup>39</sup>).

Here we report a massively parallel sequencing panel for the assessment of HPA-related SNPs useful for studies investigating the genetics that underlie behavioural and endocrine variation in aggression and the physiological stress response. For this purpose, we targeted loci of previously reported SNPs in 21 candidate genes associated with the HPA-axis. We provide detailed information on all steps from selection of target genes and polymorphisms, via laboratory work to the subsequent bioinformatics analyses of acquired massively parallel sequencing data. We additionally demonstrate the feasibility of application to faecal samples from wild populations, where non-invasive sampling is necessary.

## Materials and Methods

**Ethical statement.** The National Research Council of Thailand (NRCT) and the Department of National Parks, Wildlife and Plant Conservation (DNP) approved (permit numbers: 0004.3/3618, 0002.3/2647, 0002/17, 0002/626, 0002/2424) the data collection at the study site in Thailand and export to Germany as part of a long-term collaboration between the University of Goettingen, the German Primate Center, the DNP and Kasetsart University Bangkok under the agreement of benefit sharing. Faecal samples were collected non-invasively. No animals were harmed or sacrificed for this study. Procedures were in accordance with the American Society of Primatologists' (ASP) principles for the ethical treatment of non-human primates (<https://www.asp.org/society/resolutions/EthicalTreatmentOfNonHumanPrimates.cfm>).

**Sample collection and storage.** Samples were collected at Phu Khieo Wildlife Sanctuary in north-eastern Thailand. The sanctuary is part of the 6,500 km<sup>2</sup> protected Western Isaan Forest Complex. Faecal samples were collected from four groups of fully habituated and individually identifiable Assamese macaques. In total, we collected 478 faecal samples from 38 adult males and 41 adult females (1–15 per individual, Ø 6) over the course of the long-term field project between June 2006 and January 2016. 58% of the samples were collected between March 2015 and January 2016. Adult males are involved in reproduction, have fully developed testes and long canines. Females are considered as adult in the mating season that they first conceive in, dating back from observations of their first birth.

For genetic analyses ~5 g of faeces were collected immediately after defecation from the surface of the faecal sample from an identified individual. We applied the two-step storage procedure, which included the collection of faecal samples into 50 ml tubes (62.559.001, Sarstedt, Nümbrecht, North Rhine-Westphalia, Germany) containing 30 ml of 97% ethanol<sup>40</sup>. After storage for 24 to 36 hours, ethanol was poured off and the faecal samples were dried and stored on 30 ml silica beads (112926-00-2, Intereducation Supplies Co., Ltd., Bangkok, Thailand) in 50 ml tubes in the dark at room temperature<sup>41</sup>. These samples were exported to Germany within 6 months and then stored at –20 °C until DNA extraction was performed.

**SNP selection and amplicon primer design.** Target regions were determined by searching the literature for SNPs in candidate genes involved in stress and aggression. The majority of target regions were chosen from literature on humans (for references see Supplementary Table S1), but we also targeted macaque and pig SNPs (for full list see Supplementary Table S1). Further, we chose only target regions located in protein-coding genes that code for receptors, enzymes and transmitter molecules associated with the HPA-axis. Targets were located both in exonic and intronic as well as untranslated regions. In total, we selected 46 target regions in 21 candidate genes. A summary about all genes that were included in the multi-locus genotyping panel can be found in Table 1, for more details about the target regions, including chromosomal position, SNPs, functional consequences and selected references see Supplementary Table S1.

HGNC gene symbol	Name	Information
AVPR1B	arginine vasopressin receptor 1B	<ul style="list-style-type: none"> <li>present in the brain as well as in the pituitary where it stimulates ACTH release</li> <li>responsible for mediating the effects of vasopressin on ACTH release</li> <li>inactivation reduces aggression</li> </ul>
BDNF	brain-derived neurotrophic factor	<ul style="list-style-type: none"> <li>associated with stress vulnerability</li> <li>involved in emotion processing and cognition</li> <li>'neurotrophic hypothesis' states that stress reduces BDNF activity resulting in decreased function within limbic brain regions</li> </ul>
COMT	catechol-O-methyl transferase	<ul style="list-style-type: none"> <li>degrades catecholamines such as dopamine, epinephrine and norepinephrine</li> <li>due to its function in catecholamine degradation and dopamine inactivation it plays a pivotal role in neuro-cerebral stress processing</li> <li>essential neuropeptide for maintenance of homeostasis</li> <li>functional gene variants associated with enzyme activity, cortisol levels and aggression</li> </ul>
CRH	corticotropin-releasing hormone	<ul style="list-style-type: none"> <li>plays a crucial role in the behavioural and neuroendocrine stress response</li> <li>dysregulations are linked to stress-related psychiatric disorders</li> <li>experimental manipulations demonstrated that naturally occurring gene variants mediate individual variability in behavioural and physiological traits, thus determining an individual's coping style</li> </ul>
CRHBP	corticotropin releasing hormone binding protein	<ul style="list-style-type: none"> <li>this high-affinity protein regulates CRH availability</li> <li>widely distributed throughout the body</li> <li>variation in CRHBP expression influences the effectiveness of CRH in stimulating ACTH to release cortisol</li> </ul>
CRHR1	corticotropin-releasing hormone receptor 1	<ul style="list-style-type: none"> <li>binds CRH</li> <li>important for endocrine and behavioural stress response</li> <li>polymorphisms predict baseline cortisol and stress-related psychotic disorders</li> </ul>
CRHR2	corticotropin-releasing hormone receptor 2	<ul style="list-style-type: none"> <li>binds CRH in the membranes of hormone-sensitive cells</li> </ul>
DRD3	dopamine receptor D3	<ul style="list-style-type: none"> <li>expressed in phylogenetically old regions of the brain</li> <li>thus plays a role in cognitive and emotional functions</li> <li>allelic variants are associated with aggression and personality</li> </ul>
FKBP5	FK506 binding protein 5	<ul style="list-style-type: none"> <li>is a co-chaperone of the glucocorticoid receptor</li> <li>together they form a complex that modulates cortisol binding affinity</li> <li>thus, it modulates the HPA-axis via glucocorticoid receptor sensitivity</li> <li>gene variants associated with baseline cortisol</li> </ul>
HTR1A	serotonin receptor 1A	<ul style="list-style-type: none"> <li>binds serotonin</li> <li>activation induces the secretion of hormones including cortisol, corticosterone, ACTH and oxytocin</li> <li>important for stress-related information processing</li> <li>effects on the cortisol stress response which are explained by differences in serotonin turnover</li> <li>level of activation is associated with aggression</li> </ul>
HTR1B	serotonin receptor 1B	<ul style="list-style-type: none"> <li>widely expressed in the central nervous system</li> <li>function depends on location, in frontal cortex: postsynaptic receptor and inhibits dopamine release, in ganglia and striatum: inhibits serotonin release</li> <li>gene variants associated with vulnerability for depression and anxiety</li> </ul>
MAOA	monoamine oxidase A	<ul style="list-style-type: none"> <li>breaks down serotonin, melatonin, noradrenaline and adrenaline</li> <li>therefore, it influences synaptic concentrations of these neurotransmitters</li> <li>marker for antisocial behaviour</li> <li>gene variants linked to aggression and impulsivity in animals and humans</li> </ul>
MC2R	ACTH receptor	<ul style="list-style-type: none"> <li>binding of ACTH stimulates cortisol production</li> <li>mutations cause familial glucocorticoid deficiency</li> </ul>
NPY	neuropeptide Y	<ul style="list-style-type: none"> <li>most abundant neuropeptide in the mammalian brain and affects its activity</li> <li>plays an important role in controlling physiological processes associated with stress, especially via emotion</li> <li>released in response to stress</li> <li>referred to as a protective neurochemical that mediates stress resilience</li> </ul>
NR3C1	glucocorticoid receptor	<ul style="list-style-type: none"> <li>cortisol and other glucocorticoids bind to this receptor</li> <li>expressed in almost every cell of the body</li> <li>regulates gene transcription</li> <li>gene variants are associated with cortisol levels and psychosocial stress and can cause hypersensitivity to glucocorticoids and a poor feedback regulation of the HPA-axis</li> </ul>
OPRM1	opioid receptor mu 1	<ul style="list-style-type: none"> <li>high affinity for enkephalins and beta-endorphin</li> <li>exists mostly presynaptically</li> <li>gene variants associated with different endorphin affinities</li> <li>in macaques a variant influences HPA-axis function in response to a variety of stressors</li> </ul>
OXTR	oxytocin receptor	<ul style="list-style-type: none"> <li>presence in central nervous system</li> <li>modulates stress, anxiety, social memory and sexual, aggressive and affiliative behaviours</li> </ul>
SERPINA6	corticosteroid binding globulin	<ul style="list-style-type: none"> <li>binding and transport of glucocorticoids in mammals</li> <li>assumed that polymorphisms affect glucocorticoid transport efficiency</li> </ul>
SLC6A3	dopamine transporter	<ul style="list-style-type: none"> <li>actively removes neurotransmitters from the synaptic cleft</li> <li>thereby regulates the synaptic availability of dopamine and the duration of dopaminergic neurotransmission</li> <li>in macaques a variant is associated with dominance rank</li> </ul>
SLC6A4	serotonin transporter	<ul style="list-style-type: none"> <li>regulates the serotonin re-uptake in the synaptic cleft</li> <li>since effects of serotonin in the synapse are terminated by re-uptake, it is a crucial protein to regulate serotonin function in the brain</li> <li>functional gene variants in humans and macaques have been associated with several behavioural phenotypes</li> </ul>
TPH2	tryptophan hydroxylase 2	<ul style="list-style-type: none"> <li>it is the rate-limiting enzyme in the synthesis of serotonin</li> <li>thus, responsible for the synthesis of serotonin in brain areas</li> <li>associated with plasma cortisol</li> </ul>

**Table 1.** List of all genes included in the multi-locus genotyping panel. Note: For more details about the target regions, including chromosomal position, SNPs, functional consequences and selected references see Supplementary Table S1.

As genome data for our study species, the Assamese macaque (*Macaca assamensis*), is not available, we designed primers according to the genome sequence of the congeneric rhesus macaque (*Macaca mulatta*, v8.0.1). For detailed information about amplicon sequences of *Macaca mulatta* and human see Supplementary Table S1.

Primers were designed using the online-software Primer3Plus<sup>42</sup>. Due to DNA degradation in faecal samples, primers were designed to amplify short PCR products with a maximum of 380 bp ( $\varnothing$  207 bp), with the target SNP being as central as possible. Primer annealing temperatures were between 54 and 60 °C, with a maximum difference of 2 °C for each primer pair. Primer annealing temperature was chosen as predicted by Netprimer (Premier Biosoft, Palo Alto, California, USA). Further steps of the primer design included: (i) specificity check: Primer-Blast, NCBI, 'nr' database for Mammalia<sup>43</sup>, (ii) dimerization check: Netprimer (Premier Biosoft, Palo Alto, California, USA), (iii) secondary structure check: The mfold Web Server, DNA Folding Form<sup>44</sup>. The designed primers were ordered from Metabion (Planegg/Steinkirchen, Bavaria, Germany). In total we designed 41 primer pairs for 46 target loci (Supplementary Table S1).

**Laboratory methods.** DNA extraction was carried out with the First-DNA all-tissue Kit (D1002000, GEN-IAL GmbH, Troisdorf, North Rhine-Westphalia, Germany), following the manufacturer's protocol for DNA extraction from faeces. The protocol included an overnight incubation at 37 °C with lysis buffer 1, lysis buffer 2 and proteinase K. After centrifugation, the supernatant was combined with lysis buffer 3, incubated at –20 °C for 5 minutes and centrifuged, followed by washing steps with 70% ethanol stored at –20 °C. Finally, DNA was eluted in 50  $\mu$ l HPLC water (115333, Merck, Darmstadt, Hesse, Germany) and stored at –20 °C until further processing. All steps of the protocol were carried out with DNA LoBind Tubes (0030108051, Eppendorf AG, Hamburg, Germany). Total genomic DNA concentration was measured with a NanoDrop Spectrophotometer (ND-1000, PEQLAB Biotechnologie GmbH, Erlangen, Bavaria, Germany) and diluted to a final concentration of 100 ng/ $\mu$ l. 78% of the extracted faecal samples were collected between March 2015 and January 2016.

Target regions were amplified in 96-well plates (AB0600, Thermo Fisher Scientific, Waltham, Massachusetts, USA) with 1 U BioThermTaq DNA Polymerase (GC-002-5000, Genecraft, Cologne, North-Rhine Westphalia, Germany) in a 30  $\mu$ l PCR mix (1 x reaction buffer, 0.16 mM for each dNTP, 0.33  $\mu$ M for each primer, and 18 ng BSA, 100 ng template DNA), with the following thermocycler (Labcyler, Sensoquest, Göttingen, Lower Saxony, Germany) conditions: 2 minutes at 94 °C, 60 cycles of 30 seconds at 94 °C, 30 seconds at the appropriate annealing temperature (see Supplementary Table S1), 30 seconds at 72 °C, and 5 minutes at 72 °C. To check for PCR contamination, we ran 3 to 7 non-template controls on each 96-well plate. After amplification, aliquots were size-separated on 2% agarose gels along with a size standard (SM0241, Thermo Fisher Scientific, Waltham, Massachusetts, USA) to check for PCR performance and correct amplicon size. PCR products were then purified with Solid Phase Reverse Immobilization (SPRI) technology using 2.5x Ampure Beads (A63881, Beckman Coulter, Brea, California, USA) and again subjected to 2% agarose gel electrophoresis to control for purification performance. DNA concentration was measured with a Qubit 3.0 (Q32854, Thermo Fisher Scientific, Waltham, Massachusetts, USA). To test if our primers are target-specific, all 41 SPRI-purified amplicons of two individuals, acquired via PCR from faecal DNA extracts, were sequenced using Sanger technology. Therefore, we applied both amplification primers (3.3 pmol) and the Big Dye Cycle Sequencing Kit (433776452, Thermo Fisher Scientific, Waltham, Massachusetts, USA), and ran the reactions on an ABI 3130xl genetic analyzer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Sequence electropherograms were checked with DNA Baser (DNA Sequence Assembler v4, 2013, Heracle Biosoft S.R.L, Mioveni, Argeş, Romania) and compared with the respective target sequences of rhesus macaque and human.

For massively parallel sequencing, the amplicons from each individual were pooled in equimolar amounts to a total of 120 ng. Sequencing libraries were generated following the method described in Rohland *et al.*<sup>45</sup> without uracil-DNA-glycosylase treatment based on Meyer and Kircher<sup>46</sup> and Kircher *et al.*<sup>47</sup>. To check for performance of library preparation, we ran all libraries on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA) using the High Sensitivity DNA chip (5067-4626). Libraries were then pooled to a library mix with a final concentration of 10 nM. We added an additional step to control for quantity by running a qPCR with a 7500 Fast Real-Time PCR System (Thermo Fisher Scientific, Waltham, Massachusetts, USA) using the Bio-Rad EvaGreen Supermix (1725211, Bio-Rad, Hercules, California, USA) following the manufacturer's recommendations and using three samples of the library mix and the concentration standards (5 nM, 10 nM, 20 nM). Reactions were run under the following conditions: 2 minutes at 50 °C, 10 minutes at 95 °C, 40 cycles of 30 seconds at 95 °C, 34 seconds at 60 °C and 30 seconds at 72 °C. Sequencing was conducted on an Illumina MiSeq sequencer (paired-end 150 bp) at the Microarray and Deep-Sequencing Core Facility, University Medical Center Göttingen, Lower Saxony, Germany.

**SNP calling.** After Illumina sequencing, all produced FASTQ reads were quality-checked and trimmed with FastQC<sup>48</sup> and Trimmomatic v0.36<sup>49</sup>. For SNP calling, we used the GATK best practices pipeline for germline SNP discovery<sup>50</sup> as well as the SAMtools-bcftools-pipeline<sup>38</sup>. For both pipelines, all quality-checked reads were mapped against the genome of *Macaca mulatta* v8.0.1 using BWA MEM v0.7.12<sup>51</sup>. We followed the GATK best practice pipeline and did not mark duplicates in the bam-file, because it is not possible to distinguish duplicate reads in amplicon sequencing where the major proportion of sequences reads represents PCR duplicates<sup>52,53</sup>. A first variant call was carried out using GATK's HaplotypeCaller. Recovered SNPs were filtered using the hard-filtering procedures recommended by GATK's best practices (VariantFiltration). The following quality filter expression was applied: quality by depth smaller than 2.0 (QD < 2.0), mapping quality smaller than 40.0 (MQ < 40.0), Fisher strand (Phred-scaled p-value using Fisher's Exact Test) more than 60.0 (FS > 60.0), mapping quality rank sum (the u-based z-approximation from the Mann-Whitney Rank Sum Test for mapping qualities) smaller than –12.5 (MQRankSum < –12.5), and read position rank sum (the u-based z-approximation from the Mann-Whitney

Individuals	In percentage				In the number of individuals			
	Minimum	Maximum	Mean	Standard deviation	Minimum	Maximum	Mean	Standard deviation
homozygous for reference allele	0.00	98.73	82.60	28.00	0.00	78.00	62.65	22.88
heterozygous	0.00	56.96	10.28	15.29	0.00	45.00	7.83	11.82
homozygous for alternate allele	0.00	100.00	7.12	20.10	0.00	78.00	4.28	11.43

**Table 2.** General descriptive statistics of the detected SNPs in the investigated population of Assamese macaques.

Rank Sum Test for the distance from the end of the read for reads with the alternate allele) smaller than  $-8.0$  ( $\text{ReadPosRankSum} < -8.0$ )<sup>50</sup>. Afterwards, Base Quality Score Recalibration (BQSR) was performed twice using GATK's BaseRecalibrator and PrintReads. The final variant calling process was conducted with GATK's HaplotypeCaller in GVCF mode. The produced GVCF-files were merged using GATK's GenotypeGVCFs.

SAMtools v1.4 mpileup was also used to generate raw variant calls using *Macaca mulatta* v8.0.1 as reference genome. The following settings were used:  $-d$  (maximal per-file depth) set to 250,  $-E$  (recalculate BAQ),  $-BCF$  (generate genotype likelihoods in BCF format), and  $-output-tags$  set to DP [to get the DP (number of filtered reads covering the corresponding allele) tag in the output file]. Variant calling was done using bcftools call v1.4 with the following settings:  $-multiallelic-caller$  (alternative model for multiallelic and rare-variant calling)  $-variants-only$  (output variant sites only)  $-O v$  (output type: 'v' uncompressed VCF)<sup>38</sup>.

Only variants identified by both the GATK and the SAMtools pipeline lying within the ranges of the genes of interest were used for subsequent analyses. Additionally, we compared the number of variants called without a threshold for the Phred quality (QUAL) score, with QUAL set to be  $\geq 30$ , and with QUAL set to be  $\geq 100$ . The Phred quality score gives a logarithmically related prediction-value to the base-calling error. The higher the quality score, the higher the base call accuracy<sup>54</sup>.

**Variant annotation and effect prediction.** SnpEff v4.3i<sup>55</sup> and Variant Effect Predictor v87<sup>56</sup> (VEP) were used for variant annotation and SNP effect prediction. We compared two different applications because variant prediction software can differ in their predicted effects<sup>57</sup>. With the VEP plugin 'MaxEntScan' it was possible to compare scores of the reference and mutant splice sites using a maximum entropy model and to predict splice site effects. Additionally, linkage disequilibrium scores were calculated using vcftools v0.1.14<sup>58</sup> with the option 'geno-r2'. Subsequently, all calculated annotations and effects were analysed in detail by hand.

## Results

Sanger sequencing of the 41 amplicons from two macaque individuals revealed that the applied primers are target-specific. From the 79 investigated macaque individuals, we obtained a total of 3066 amplicons, with a minimum of 34 and a maximum of 41 amplicons per individual ( $\bar{O}$  38.6). For 10 of the 79 individuals, the first faecal DNA extracts (3 samples collected in 2007 and 7 samples collected in 2015) did not yield any amplicons (probably due to inhibitors in faeces), however, amplicons were successfully obtained from second extracts derived from other samples from the same individuals in all cases.

After sequencing on Illumina's MiSeq, we obtained a total of 23,604,930 reads. Around 95% of the reads exhibit a Phred score  $> 29$ . About 85% of the reads could be mapped against the reference genome of *Macaca mulatta* v8.0.1 with a mean depth of 3,219.81 and a mean mapping quality of 59.02. Detailed sequencing statistics can be found in Supplementary Tables S2 and S3.

**Variant calling.** We compared different variant calling approaches in order to use only those variants reproduced by multiple pipelines. Using the complete dataset of shared variants without consideration of any filtering steps (QUAL, QD, MQ, FS, MQRankSum, and ReadPosRankSum), 70.12% of the variants called by SAMtools were also found with the GATK pipeline, whereas only 28.31% of the variants called by GATK were also found by SAMtools (Supplementary Figure S1). In sum, SAMtools called 230 SNPs, whereas GATK detected 373 SNPs. In total, 169 SNPs in 21 genes were verified using both variant calling approaches. 10 out of the 169 SNPs showed homozygosity for the alternate allele and thus represented simply a difference to the rhesus macaque genome, whereas 159 SNPs were identified as polymorphic sites within the study population. General descriptive statistics of the distribution of individuals being homozygous for the reference or alternate allele, or heterozygous, can be found in Table 2.

Using the described filtering steps without consideration of the QUAL scores, the number of detected variants by GATK decreased to 170 SNPs resulting in 144 SNPs in 20 genes shared by both callers. Using different QUAL scores as further selection steps, the number of detected polymorphisms changed again. Extracting only those variants with a QUAL score  $\geq 30$ , GATK still detected 170 SNPs, but SAMtools called only 194 SNPs (Supplementary Figure S1). Only 140 SNPs in 20 genes were shared by both callers. With a QUAL score  $\geq 100$ , the number of variants decreased further to 124 shared SNPs in 20 genes out of 170 SNPs detected with GATK and 162 SNPs detected with SAMtools (Supplementary Figure S1).

**Variant effect prediction.** The variants detected by both the SAMtools and the GATK pipeline were used to predict possible variant effects. For this purpose, we used two applications: VEP and SnpEff. Comparing the results, almost all variants were classified as the same consequence type in both applications. The only differences

Consequence type	VEP		SnpEff	
	Count	Percent	Count	Percent
3_prime_UTR_variant	13	2.36%	13	2.23%
5_prime_UTR_premature_start_codon_gain_variant	0	0%	2	0.34%
5_prime_UTR_variant	6	1.09%	6	1.03%
downstream_gene_variant	20	3.62%	20	3.44%
frameshift_variant	0	0.00%	0	0.00%
intergenic_variant	11	1.99%	39	6.70%
intron_variant	260	47.10%	260	44.67%
missense_variant	62	11.23%	62	10.65%
splice_region_variant	19	3.44%	19	3.26%
synonymous_variant	105	19.02%	105	18.04%
upstream_gene_variant	56	10.14%	56	9.62%
total	552	100.00%	582	100.00%

**Table 3.** Count and percent of consequence types of SNPs predicted by VEP and SnpEff.

Consequence type	VEP		SnpEff	
	Count	Percent	Count	Percent
moderate	122	22.89%	124	22.03%
low	62	11.63%	62	11.01%
modifier	349	65.48%	377	66.96%
total	533	100%	563	100.00%

**Table 4.** Count and percent of impact classes of SNPs predicted by VEP and SnpEff.

are two counts of ‘5’-UTR premature start codon gain variant’ found with SnpEff, but not with VEP and 25 additional ‘intergenic variants’ found by SnpEff (Table 3). Consequently, the predicted variant effects show a high similarity between both methods concerning their impact (Table 4).

Using the VEP plugin MaxEntScan, two SNPs were found to be associated with different entropies at splice sites. These SNPs were located in the serotonin transporter (*SLC6A3*) and the neuropeptide Y (*NPY*) genes and changed entropy from 0.40 to  $-4.66$  and  $9.38$  to  $9.43$ , respectively. These SNPs were identified as ‘splice\_region\_variant’ and ‘intron\_variant’ by VEP and SnpEff. However, not all identified ‘splice\_region\_variants’ caused differences in splice site entropy, as predicted by MaxEntScan. An analysis of linkage disequilibrium using vcftools revealed that 64 SNPs were linked. Linkage  $r^2$  values ranged from 0.28 to 1 ( $\emptyset$  0.91).

## Discussion

Enhancing the approach of behavioural genetics and physiogenetics in wild animals would extend our knowledge of the factors that contribute to the still not completely understood variation within and between populations under natural selection pressures<sup>8</sup>. Behavioural studies often investigate the impacts of personality, age, sex or external factors such as social environment, group size and dominance hierarchy to explain inter-individual differences in HPA-axis related traits and probably misrepresent the amount of variation to be explained by such factors because they neglect genetic impacts<sup>59–62</sup> or focus only on one or two gene variants<sup>31–33</sup>, but see<sup>34,63,64</sup>. An extended integration of genotype information in wild populations will facilitate a more comprehensive understanding of the observed phenotypic variation. For example, variant information on multiple loci can be used to generate cumulative genetic risk scores to predict individual variation<sup>35,65</sup>. Among the best-studied aspects of animal behaviour in the wild are behavioural and physiological reactions to social and environmental stressors<sup>66</sup>. Behavioural and physiological responses are tightly linked to the HPA-axis, the main physiological pathway activated in response to stressful stimuli<sup>67</sup>. Thus, genes coding for the components of the HPA-axis, which act in concert to maintain homeostatic balance, are important targets for the investigation of phenotypic variation in stress- and aggression-related traits.

We provide a SNP panel that may serve as a basic tool for future studies investigating the genetics of stress and aggression in behavioural and ecological studies. The offered panel and protocol enables field biologists teaming up with a laboratory to screen entire wild animal populations for multiple highly interesting target regions in a fast and parallel way. This study demonstrates that polymorphisms at purportedly functional sites exist in HPA-linked genes in natural populations. Knowing the samples from individually-recognized individuals, which is usually the case in long-term studies on wild animal groups, allows to accurately determine population frequencies of SNPs. Further, this application is transferable to other species. The HPA-axis is conserved among vertebrates<sup>10,11</sup> and orthologous gene regions can be found easily for the species of interest e.g. via BLAST<sup>43</sup>, PSI-BLAST<sup>68</sup>, BLAT<sup>69</sup>, SSEARCH<sup>70</sup> (biology.wustl.edu/gcg/ssearch.html) or HMMER3 (hmmer.org) searches. For application to other species, we recommend searching for the primer or amplicon sequences (Supplementary Table S1) using the aforementioned software and choosing the respective sequences to design taxon-specific

primers. All subsequent steps can be carried out as outlined in our protocol. With small PCR product sizes, as in our case, allelic dropout is a marginal problem, but to further minimize the risk of allelic dropout we recommend multiple PCRs per sample<sup>71</sup> or to perform replicates from different samples of the same individual.

Given that acquired massively parallel sequencing data hold many opportunities for further in-depth analyses, we provide detailed information on the applied bioinformatics pipelines. For example, with the help of GATK<sup>37</sup> or SAMtools<sup>38</sup> variants can be detected and used for subsequent high-throughput analyses concerning their functionality and possible effects (e.g. VEP<sup>56</sup> and SnpEff<sup>55</sup>). However, our analyses revealed that GATK called more variants than SAMtools in all conditions of the different quality thresholds and emphasize the importance of comparing pipelines and relying on validated, intersecting sets of SNPs. Further analyses of, e.g. splice site entropy (MaxEntScan) and linkage disequilibrium (vcftools), help to interpret the effects of detected polymorphisms and their potential consequences on physiological pathways.

To promote this application for studies on wild populations of non-model organisms, in which the consideration of genotype is particularly rare, we established our methods based on faecal samples. Studies on protected, free-ranging animals are often confined to the non-invasive collection of genetic material. DNA extracts from such low-quality sources contain only small amounts of host DNA<sup>72</sup>. The dominance of exogenous, non-host, e.g. microbiome or food DNA, rules out a massively parallel sequencing-application on the pure DNA extracts, without prior amplification or enrichment of target regions. Sequencing of amplicons with the traditional Sanger method is time- and cost-intensive, particularly when encountering larger sets of target regions and individuals (3066 amplicons in this study). Furthermore, when two SNPs are found in one amplicon, the haplotype structure remains unknown. When applying the classical Sanger sequencing approach, elucidating haplotypes requires additional working steps, such as cloning. Here we have shown that multiple target regions can be easily covered with massively parallel amplicon sequencing from faecal DNA of larger numbers of individuals. Alternatively, target regions could be captured using synthesized or self-made capture probes<sup>73,74</sup>. While such methods may reach better sequencing uniformity, they are less target-specific and exhibit lower average coverage than amplicon-based technologies<sup>75</sup>. However, such methods could be applied to calculate additional background population structure<sup>74</sup>. Another important aspect, especially for studies on wild populations that are often limited to low-quality DNA samples, is that amplicon-based massively parallel sequencing methods allow processing of low-input DNA samples<sup>75</sup>. Further, due to the large amount of sequence reads produced, the regions of interest (amplicons) exhibit high coverage reducing the detection of false positive variants. A caveat of the study is that it is still ultimately a bottom-up approach needing a priori information to select target regions. As technology will improve in the future, top-down approaches will most likely also become an effective and economical tool for low-quality samples making more data available. These top-down approaches could be applied to generate haplotype data for a multitude of loci across the genome in a fast and parallel way, to calculate relatedness and include kinship relations in behavioural genetics approaches as well as to identify conserved genome regions or gene segments with high mutation rates in the investigated populations.

Numerous field studies have established links between non-invasive measures of HPA-axis activity and the behaviour of animals. Glucocorticoid metabolite levels, the end products of the HPA-axis, increase during reproductive challenges<sup>19,76</sup>, with increasing aggression given or received<sup>76,77</sup> and are often related to social status<sup>78</sup>. Mostly lab-based studies have established links between HPA-axis activity and genetic variation at individual loci. For example, a mu-opioid receptor polymorphism is associated with cortisol and aggressive threat scores<sup>79</sup> and variation in the serotonin transporter gene is associated with increased HPA-axis activity<sup>80</sup> in captive primates. Progress is hampered by a lack of (1) integration of both research streams to link genetic variation to HPA-axis activity and behaviour, and (2) studies screening multiple loci involved in HPA-axis regulation at the same time. We propose that our panel can serve as a basis for general behavioural studies aiming to extend their study design on a molecular level and step into the field of behavioural genetics. The simultaneous investigation of genes and behaviour will help to achieve a more comprehensive understanding of individual animal characteristics.

## Data Accessibility

Massively parallel sequencing-data were submitted to the Sequence Read Archive (SRA) available via NCBI with the accession number SRP116685. Additional data are available in the Supplementary Information.

## References

- Ekblom, R. & Galindo, J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*. **107**, 1–15 (2011).
- Perry, G. H. The promise and practicality of population genomics research with endangered species. *Int. J. Primatol.* **35**, 55–70 (2014).
- Craig, I. W. & Halton, K. E. Genetics of human aggressive behaviour. *Hum. Genet.* **126**, 101–113 (2009).
- Pavlov, K. A., Chistiakov, D. A. & Chekhonin, V. P. Genetic determinants of aggression and impulsivity in humans. *J. Appl. Genet.* **53**, 61–82 (2012).
- Plomin, R., DeFries, J. C., Knopik, V. S. & Neiderhiser, J. *Behavioral Genetics* (Worth Publishers, 2013).
- Linkowski, P. *et al.* Twin study of the 24-h cortisol profile: Evidence for genetic control of the human circadian clock. *Am. J. Physiol. Endocrinol. Metab.* **264**, E173–E181 (1993).
- Miles, D. R. & Carey, G. Genetic and environmental architecture on human aggression. *J. Pers. Soc. Psychol.* **72**, 207–217 (1997).
- Brent, L. J. N. & Melin, A. D. The genetic basis of primate behavior: Genetics and genomics in field-based primatology. *Int. J. Primatol.* **35**, 1–10 (2014).
- Tung, J., Alberts, S. C. & Wray, G. A. Evolutionary genetics in wild primates: Combining genetic approaches with field studies of natural populations. *Trends Genet.* **26**, 353–362 (2010).
- Denver, R. J. Structural and functional evolution of vertebrate neuroendocrine stress systems. *Ann. NY Acad. Sci.* **1163**, 1–16 (2009).
- Schulkin, J. Evolutionary conservation of glucocorticoids and corticotropin releasing hormone: behavioral and physiological adaptations. *Brain Res.* **1392**, 27–46 (2011).

12. Øverli, Ø., Kotzian, S. & Winberg, S. Effects of cortisol on aggression and locomotor activity in rainbow trout. *Horm. Behav.* **42**, 53–61 (2002).
13. Carere, C., Groothuis, T. G. G., Möstl, E., Daan, S. & Koolhaas, J. M. Fecal corticosteroids in a territorial bird selected for different personalities: Daily rhythm and the response to social stress. *Horm. Behav.* **43**, 540–548 (2003).
14. Huhman, K. L., Moore, T. O., Ferris, C. F., Mougey, E. H. & Meyerhoff, J. L. Acute and repeated exposure to social conflict in male golden hamsters: increases in plasma POMC-peptides and cortisol and decreases in plasma testosterone. *Horm. Behav.* **25**, 206–216 (1991).
15. Salas, M. *et al.* Aggressive behavior and hair cortisol levels in captive Dorcas gazelles (*Gazella dorcas*) as animal-based welfare indicators. *Zoo Biol.* **35**, 467–473 (2016).
16. Finkler, H. & Terkel, J. Cortisol levels and aggression in neutered and intact free-roaming female cats living in urban social groups. *Physiol. Behav.* **99**, 343–347 (2010).
17. Rosado, B. *et al.* Blood concentrations of serotonin, cortisol and dehydroepiandrosterone in aggressive dogs. *Appl. Anim. Behav. Sci.* **123**, 124–130 (2010).
18. Honess, P. E. & Marin, C. M. Behavioural and physiological aspects of stress and aggression in nonhuman primates. *Neurosci. Biobehav. Rev.* **30**, 390–412 (2006).
19. Ostner, J., Kappeler, P. & Heistermann, M. Androgen and glucocorticoid levels reflect seasonally occurring social challenges in male redfronted lemurs (*Eulemur fulvus rufus*). *Behav. Ecol. Sociobiol.* **62**, 627–638 (2008).
20. Halász, J., Liposits, Z., Kruk, M. R. & Haller, J. Neural background of glucocorticoid dysfunction-induced abnormal aggression in rats: Involvement of fear- and stress-related structures. *Eur. J. Neurosci.* **15**, 561–569 (2002).
21. Kruk, M. R., Halász, J., Meelis, W. & Haller, J. Fast positive feedback between the adrenocortical stress response and a brain mechanism involved in aggressive behavior. *Behav. Neurosci.* **118**, 1062–1070 (2004).
22. Soma, K. K., Scotti, M. A. L., Newman, A. E. M., Charlier, T. D. & Demas, G. E. Novel mechanisms for neuroendocrine regulation of aggression. *Front. Neuroendocrinol.* **29**, 476–489 (2008).
23. Jabbi, M. *et al.* Convergent genetic modulation of the endocrine stress response involves polymorphic variations of 5-HTT, COMT and MAOA. *Mol. Psychiatry.* **12**, 483–490 (2007).
24. Zhou, Z. *et al.* Genetic variation in human NPY expression affects stress response and emotion. *Nature.* **452**, 997–1001 (2008).
25. Qayyum, A. *et al.* The role of the catechol-o-methyltransferase (COMT) GeneVal158Met in aggressive behavior, a review of genetic studies. *Curr. Neuropharmacol.* **13**, 802–814 (2015).
26. Craig, I. W. The importance of stress and genetic variation in human aggression. *Bioessays* **29**, 227–236 (2007).
27. D'souza, U. M. & Craig, I. W. Functional polymorphisms in dopamine and serotonin pathway genes. *Hum. Mutat.* **27**, 1–13 (2006).
28. Kormos, V. & Gaszner, B. Role of neuropeptides in anxiety, stress, and depression: From animals to humans. *Neuropeptides.* **47**, 401–419 (2013).
29. Molinoff, P. B. & Axelrod, J. Biochemistry of catecholamines. *Annu. Rev. Biochem.* **40**, 465–400 (1971).
30. Subbannaya, T. *et al.* An integrated map of corticotropin-releasing hormone signaling pathway. *J. Cell Commun. Signal.* **7**, 295–300 (2013).
31. Garamszegi, L. Z. *et al.* The relationship between DRD4 polymorphisms and phenotypic correlations of behaviors in the collared flycatcher. *Ecol. Evol.* **4**, 1466–1479 (2014).
32. Kalbitzer, U. *et al.* Insights into the genetic foundation of aggression in *Papio* and the evolution of two length-polymorphisms in the promoter regions of serotonin-related genes (5-HTTLPR and MAOALPR) in Papionini. *BMC Evol. Biol.* **16**, 121, <https://doi.org/10.1186/s12862-016-0693-1> (2016).
33. Timm, K., Tilgar, V. & Saag, P. DRD4 gene polymorphism in great tits: gender-specific association with behavioural variation in the wild. *Behav. Ecol. Sociobiol.* **69**, 729–735 (2015).
34. Bergey, C. M., Phillips-Conroy, J. E., Disotell, T. R. & Jolly, C. J. Dopamine pathway is highly diverged in primate species that differ markedly in social behavior. *PNAS.* **113**, 6178–6181 (2016).
35. Ferguson, B. *et al.* Genetic load is associated with hypothalamic–pituitary–adrenal axis dysregulation in macaques. *Genes Brain Behav.* **11**, 949–957 (2012).
36. Pearce, E., Wlodarski, R., Machin, A. & Dunbar, R. I. M. Variation in the  $\beta$ -endorphin, oxytocin, and dopamine receptor genes is associated with different dimensions of human sociality. *PNAS.* **114**, 5300–5305 (2017).
37. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
38. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics.* **25**, 2078–2079 (2009).
39. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
40. Nsubuga, A. M. *et al.* Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Mol. Ecol.* **13**, 2089–2094 (2004).
41. Roeder, A. D., Archer, F. I., Poinar, H. N. & Morin, P. A. A novel method for collection and preservation of faeces for genetic studies. *Mol. Ecol. Notes.* **4**, 761–764 (2004).
42. Untergasser, A. *et al.* Primer3 - new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115, <https://doi.org/10.1093/nar/gks596> (2012).
43. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
44. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
45. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624, <https://doi.org/10.1098/rstb.2013.0624> (2015).
46. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **6**, pdb-prot5448, <https://doi.org/10.1101/pdb.prot5448> (2010).
47. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3, <https://doi.org/10.1093/nar/gkr771> (2012).
48. Andrews, S. FastQC: A quality control tool for high throughput sequence data. *Babraham Institute* <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* **30**, 2114–2120 (2014).
50. Van der Auwera, G. A. *et al.* From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics.* **43**, 1–33 (2013).
51. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at, <https://arxiv.org/abs/1303.3997> (2013).
52. Tata, P. *et al.* Biologic Basis of Personalized Therapy in Head and Neck Squamous Cell Carcinoma. In *Contemporary Oral Oncology: Biology, Epidemiology, Etiology, and Prevention* (ed. Kuriakose, M. A.) 461–486 (Springer, 2017).
53. Ebbert, M. T. *et al.* Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinformatics.* **17**, 239 (2016).
54. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).

55. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118, iso-2, iso-3. *Fly*. **6**, 80–92 (2012).
56. McLaren, W. *et al.* The ensembl variant effect predictor. *Genome Biol.* **17**, 122, <https://doi.org/10.1186/s13059-016-0974-4> (2016).
57. Wertz, J., Bair, T. B. & Chimenti, M. S. PyVar: An extensible framework for variant annotator comparison. Preprint at, <https://www.biorxiv.org/content/early/2016/09/30/078386> (2016).
58. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
59. Seyfarth, R. M., Silk, J. B. & Cheney, D. L. Variation in personality and fitness in wild female baboons. *PNAS*. **109**, 16980–16985 (2012).
60. Young, C., Majolo, B., Heistermann, M., Schülke, O. & Ostner, J. Responses to social and environmental stress are attenuated by strong male bonds in wild macaques. *PNAS*. **111**, 18195–18200 (2014).
61. Baugh, A. T., Davidson, S. C., Hau, M. & van Oers, K. Temporal dynamics of the HPA axis linked to exploratory behavior in a wild European songbird (*Parus major*). *Gen. Comp. Endocrinol.* **250**, 104–112 (2017).
62. Mell, H. *et al.* Do personalities co-vary with metabolic expenditure and glucocorticoid stress response in adult lizards? *Behav. Ecol. Sociobiol.* **70**, 951–961 (2016).
63. Song, S. *et al.* Targeted next-generation sequencing for identifying genes related to horse temperament. *Genes & Genomics*. **39**, 1325–1333 (2017).
64. Madlon-Kay, S. *et al.* Weak effects of common genetic variation in oxytocin and vasopressin receptor genes on rhesus macaque social behavior. *Am. J. Primatol.* e22873, <https://doi.org/10.1002/ajp.22873> (2018).
65. Belsky, D. W. & Israel, S. Integrating genetics and social science: Genetic risk scores. *Biodemography Soc. Biol.* **60**, 137–155 (2014).
66. Reeder, D. M. & Kramer, K. M. Stress in free-ranging mammals: Integrating physiology, ecology, and natural history. *J. Mammal.* **86**, 225–235 (2005).
67. Del Rey, A., Chrousos, G. & Besedovsky, H. *The hypothalamus-pituitary adrenal axis.* (Elsevier, 2008).
68. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
69. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
70. Smith, T. F. & Waterman, M. S. Comparison of biosequences. *Adv. Appl. Math.* **2**, 482–489 (1981).
71. Goodrich, J. K. *et al.* Conducting a microbiome study. *Cell*. **158**, 250–262 (2014).
72. Perry, G. H., Marioni, J. C., Melsted, P. & Gilad, Y. Genomic-scale capture and sequencing of endogenous DNA from feces. *Mol. Ecol.* **19**, 5332–5344 (2010).
73. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*. **5**, e14004, <https://doi.org/10.1371/journal.pone.0014004> (2010).
74. Snyder-Mackler, N. *et al.* Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. *Genet.* **203**, 699–714 (2016).
75. Samorodnitsky, E. *et al.* Evaluation of hybridization capture versus amplicon-based methods for whole-exome sequencing. *Hum. Mutat.* **36**, 903–914 (2015).
76. Ostner, J., Heistermann, M. & Schülke, O. Dominance, aggression and physiological stress in wild male Assamese macaques (*Macaca assamensis*). *Horm. Behav.* **54**, 613–619 (2008).
77. Wittig, R. M., Crockford, C., Weltring, A., Deschner, T. & Zuberbühler, K. Single aggressive interactions increase urinary glucocorticoid levels in wild male chimpanzees. *PLoS One*. **10**, e0118695, <https://doi.org/10.1371/journal.pone.0118695> (2015).
78. Goymann, W. & Wingfield, J. C. Allostatic load, social status and stress hormones: The costs of social status matter. *Anim. Behav.* **67**, 591–602 (2004).
79. Miller, G. M. *et al.* A mu-opioid receptor single nucleotide polymorphism in rhesus monkey: association with stress response and aggression. *Mol. Psychiatry*. **9**, 99–108 (2004).
80. Barr, C. S. *et al.* Rearing condition and rh5-HTTLPR interact to influence limbic-hypothalamic-pituitary-adrenal axis response to stress in infant macaques. *Biol. Psychiatry*. **55**, 733–738 (2004).

## Acknowledgements

We thank the National Research Council of Thailand (NRCT) and the Department of National Parks, Wildlife and Plant Conservation (DNP) for permission to conduct this study and for all the support granted. We are grateful to J. Prabnasuk, K. Nitaya, T. Wongsanak, M. Pongjantarasiem and K. Kreetiyutanont, M. Kumsuk, W. Saenphala (Phu Khieo Wildlife Sanctuary) for their cooperation and permission to carry out this study. We thank A. Koenig and C. Borries, who developed the field site at Huai Mai Sot Yai. Special thanks go to S. Jumrudwong, W. Nueorngshiyos, N. Juntuch, J. Wanart, R. Intalo, T. Kilawit, N. Pongangan, B. Klaewklar, N. Bualeng, A. Ebenau, P. Saisawatdikul, K. Srithorn, M. Swagemakers, and T. Wisate for their excellent help in the field. We acknowledge C. Schwarz, N. Westphal, F. Aron and L. Walter for their support in the genetics lab. We thank our colleagues, particularly F. Ludewig, from the Microarray and Deep-Sequencing Core Facility, University Medical Center Göttingen for sequencing. We also thank T. S. Gutleb for valuable comments on the manuscript. We thank the members of the research unit ‘Sociality and Health in Primates’ (DFG FOR 2136) for stimulating discussions. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number OS201/8-1; 254142454/GRK 2070.

## Author Contributions

D.R.G., J.O., O.S. and C.R. designed research, D.R.G., J.O., O.S., W.W. and M.S. collected data in the field, D.R.G. performed research, A.N. performed bioinformatics analyses, D.R.G. and A.N. analysed data, D.R.G. wrote the paper, J.O., O.S., W.W., M.S., C.R. and A.N. commented on the paper.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-34223-y>.

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018