GENOMIC RESOURCES NOTE
Genomic Resources Notes Accepted 1 December 2014 – 31 January 2015

GENOMIC RESOURCES DEVELOPMENT CONSORTIUM,1 SIMON BLANCHET,2,3 OLIVIER BOUCHEZ,4,5 COLIN A. CHAPMAN,6,7 PAUL D. ETTER,8 TONY L. GOLDBERG,9 ERIC A. JOHNSON,8 JAMES H. JONES,10 GÉRALDINE LOOT,2,3 PATRICK A. OMEJA,11 OLIVIER REY,2 MARIA JOSE RUIZ-LOPEZ,12 WILLIAM M. SWITZER13 and NELSON TING12,14

1Molecular Ecology Resources Editorial Office, 6270 University Blvd, Vancouver, British Columbia V6T 1Z4, Canada, 2Station d’Ecologie Expérimentale du CNRS à Moulis, USR 2936, 09 200 Moulis, France, 3Université de Toulouse, UPS, UMR15174 (EDB), 118 route de Narbonne, 31062 Toulouse cedex 9, France, 4GeT-PlaGe, Genotoul, INRA, Auzeville, F31326 Castanet-Tolosan, France, 5INRA, UMR1388 Génétique, Physiologie et Systèmes d’Elevage, Elevage, F31326 Castanet-Tolosan, France, 6Department of Anthropology & McGill School of Environment, McGill University, Montreal, Quebec H3A 2T7, Canada, 7Wildlife Conservation Society, 2300 Southern Boulevard, Bronx, NY 10460, USA, 8Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA, 9School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA, 10Department of Anthropology, Stanford University, Stanford, CA 94305, USA, 11Makerere University Biological Field Station, P.O Box 967, Fort Portal, Uganda, 12Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403, USA, 13Laboratory Branch, Division of HIV/AIDS Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia 30329, USA, 14Department of Anthropology, University of Oregon, Eugene, OR 97403, USA

Abstract

This article documents the public availability of (i) transcriptome sequence data and assembly for the rostrum dace (Leuciscus burdigalensis) naturally infected by a copepod ectoparasite (Tracheliastes polycolpus) and (ii) SNPs identified and validated from RAD sequencing for the Ugandan red colobus (Procolobus rufomitratus tephrosceles) using RAD sequencing.

Table 1 contains information on the focal species, data type and resource developed, as well as access details for the data. The authors responsible for each genomic resource are listed in the final column. Full descriptions of how each resource was developed and tested are uploaded as Supplemental Information with the online version of this manuscript.

Table 1 Information on the focal species, data type and resource developed, as well as access details for the data. The authors responsible for each genomic resource are listed in the final column

<table>
<thead>
<tr>
<th>Species (no. of individuals)</th>
<th>Data type</th>
<th>Resources</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procolobus rufomitratus tephrosceles (24)</td>
<td>SNP discovery and validation from RAD sequencing</td>
<td>RAD sequence data: NCBI Sequence Read Archive SRP050135 SNP information including validation results: <a href="http://dx.doi.org/10.5061/dryad.404fh">http://dx.doi.org/10.5061/dryad.404fh</a></td>
<td>Maria Jose Ruiz-Lopez, Tony L. Goldberg, Colin A. Chapman, Patrick A. Omeja, James H. Jones, William M. Switzer, Paul D. Etter, Eric A. Johnson, Nelson Ting</td>
</tr>
</tbody>
</table>

Correspondence: Genomic Resources Development Consortium, E-mail: editorial.office@molecol.com

Identification of SNP markers for the endangered Ugandan red colobus (*Procolobus rufomitratus tephrosceles*) using RAD sequencing

Authors
Maria Jose Ruiz-Lopez¹, Tony L. Goldberg², Colin A. Chapman³, Patrick A. Omeja⁴, James H. Jones⁵, William M. Switzer ⁶, Paul D. Etter⁷, Eric A. Johnson⁷, Nelson Ting¹,⁸

Affiliations
¹Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403, USA
²School of Veterinary Medicine, University of Wisconsin-Madison, WI 53706, USA
³Department of Anthropology  & McGill School of Environment, McGill University, Montreal, Quebec, Canada, H3A 2T7, Wildlife Conservation Society, 2300 Southern Boulevard, Bronx, New York, USA 10460,
⁴Makerere University Biological Field Station, P.O Box 967, Fort Portal, Uganda.
⁵Department of Anthropology, Stanford University, Stanford, CA 94305, USA
⁶Laboratory Branch, Division of HIV/AIDS Prevention. National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia 30329, USA
⁷Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA
⁸Department of Anthropology, University of Oregon, Eugene, OR 97403, USA

Keywords
Red colobus, SNPs, RAD-Seq, endangered, evolution, conservation genomics
Introduction

Recent progress in DNA sequencing technologies coupled with rapid decreases in per-nucleotide sequencing costs have facilitated the discovery of large numbers of single nucleotide polymorphism (SNPs) both in model and non-model species (Ekblom & Galindo 2011). However, the discovery of SNPs in non-model species is still challenging due to the lack of reference genomes and the computational complexity of whole genome de novo assembly. “Reduced representation” approaches that reduce the genome complexity have thus grown in popularity. Among these approaches, Restriction-site-associated DNA sequencing (RAD-Seq; Baird et al. 2008; Davey et al. 2011) has proven to be a reliable method, enabling direct sequencing of homologous regions evenly distributed throughout the genome and allowing the identification and typing of thousands of SNPs. RAD-Seq has been successfully used in a number of molecular ecology studies, including research in gene flow, population structure, phylogeography, phylogenetics, and adaptation (e.g., Catchen et al. 2013a; Hohenlohe et al. 2011; Stolting et al. 2013; White et al. 2013).

Despite the availability of genomic resources across a broad array of primate taxa, none are currently available for any leaf-eating monkey (subfamily Colobinae). Furthermore, although RAD-Seq has been used broadly, it has rarely been used in primates (but see Bergey et al. 2013 and Evans et al. 2014). Here, we used RAD-Seq to discover genome-wide SNPs for a red colobus monkey (genus Procolobus; subgenus Piliocolobus) (Grubb et al. 2003; Oates & Davies 1994). These are tropical forest-adapted, arboreal, leaf-eating monkeys distributed across Equatorial Africa in 18 different morphotypes. They are among the most threatened of African primates (Oates et al. 2008; Struhsaker 2005; Ting 2008a) and are considered a good indicator species because of their sensitivity to habitat loss and hunting by humans (Mittermeier et al. 2009; Struhsaker 2005). In addition, they are an emerging model for the study of infectious disease discovery and transmission (e.g., Goldberg et al. 2009; Lauck et al. 2013; Paige et al. 2014). Currently, genetic resources for red colobus (and the colobine subfamily in general) are limited to mitochondrial DNA and microsatellites, which are typically cross-amplified using primers specifically designed for humans (Allen et al.
2012; Ting 2008b). Thus, a red colobus genomic resource will assist in conservation
efforts by allowing a better understanding of neutral and adaptive genetic diversity,
patterns of gene flow, potential adaptation to known pathogens, and how best to
construct informed management plans.

We developed a SNP panel for the Ugandan red colobus (*P. rufomitratus tephrosceles*),
focusing on the largest viable population of this endangered taxon (Kibale National
Park; Chapman *et al.* 2010; Struhsaker 2005). We analyzed 24 individuals using single-
end RAD-Seq and obtained 70,773,857 total reads, of which 58,814,906 passed the
filtering steps. Using a *de novo* pipeline we identified a total of 173,625 putative RAD
loci. To minimize potential drop out we followed the protocol of Davey *et al.* (2013) and
selected loci that were present in 90% of individuals. We found 113,376 loci, of which
50,558 were polymorphic. To our knowledge this SNP dataset represents the largest
genomic resource for the red colobus available to date.

**Data Access**
- Red colobus sequence files (fastq) for the 24 individuals analyzed have been made
available through the Sequence Read Archive (SRA) repository at NCBI (SRP050135)
- Information on the RAD loci identified is available through Dryad
  ([http://dx.doi.org/10.5061/dryad.404fh](http://dx.doi.org/10.5061/dryad.404fh))

**Meta Information**

- **Sequencing center:** University of Oregon Genomics Core Facility (Eugene,
  Oregon, USA).
- **Platform and model:** Illumina HiSeq™ 2000
- **Design description:** Our goal was to obtain a genome-wide panel of SNPs in the
  Ugandan red colobus monkey. This panel will be the first genomic resource for
  the species and will facilitate further conservation, ecological, and evolutionary
  studies. We analyzed blood samples from 24 individuals of a single population at
Kibale National Park (0° 33' N and 30° 21' E) and prepared a single-end RAD-Seq library and analyzed the data using the program STACKs v. 1.1 (available at http://creskolab.uoregon.edu/stacks/; Catchen et al. 2013b; Catchen et al. 2011).

- **Analysis type**: Restriction site associated DNA (RAD) sequencing.
- **Run date**: December 2013

**Library**

- **Strategy**: Single-end RAD-Sequencing
- **Taxon**: Procolobus rufomitratus tephrosceles
- **Sex**: Details in Table 2
- **Tissue**: blood
- **Location**: We collected samples from 2 social groups (Table 2) part of a single population at Kibale National Park, Uganda.
- **Sample handling**: 22 EDTA-treated whole blood samples were collected in 2006 from anesthetized individuals. Plasma was separated in the field by centrifugation and plasma and whole blood samples were stored in liquid nitrogen for transport. Two additional samples were collected in 2010 (Table 1).
- **Additional sample information**: Additional details on sample collection methods can be found in Goldberg et al. (2009).
- **Selection**: Reduced representation library using restriction enzyme cutting (SbfI-HF).
- **Layout**: Single end 1x150bp sequencing, V3 chemistry.
- **Library Construction Protocol**: Genomic DNA was extracted from blood samples using QIAamp DNA Mini Kits (Qiagen, Inc., Valencia, California, USA) following the manufacturers protocol. All samples were quantified using the Qubit™ 2.0 High-Sensitivity dsDNA Assay (Life Technologies, Inc., Carlsbad, California, USA) and the mean fragment size of each sample was further assessed in a Fragment Analyzer (Advanced Analytical Technologies, Inc., Ames, Iowa, USA). DNA from the 24 individuals was sequenced as a part of a larger single-end library that included another primate species and a total of 46 monkeys. For
preparing the library. 500 ng of genomic DNA from each sample was digested for 60 min at 37°C in a 50 µl reaction volume containing 5 µl 10x CutSmart™ Buffer and 10 units (U) SbfI-HF (New England Biolabs [NEB], Inc., Ipswich, Massachusetts, USA). Each sample was heat inactivated for 20 min at 65°C and allowed to cool at room temperature (RT) overnight. 4 µl of 6 base pair (bp) barcoded SbfI-P1 Adapters (100 nM) were added to each sample followed by 1.0 µl of 10x NEB Buffer 2, 0.6 µl rATP (100 mM, Promega Corporation, Madison, Wisconsin, USA), 0.5 µl (1000 U) T4 DNA ligase (2 million U/ml, NEB), 3.9 µl H₂O and incubated at RT for 30 min. Samples were heat inactivated for 20 min at 65°C and allowed to cool at RT for one hour. 30 µl of the ligation reaction for each sample was combined in sub-libraries according to fragment size. Each pooled sub-library was randomly sheared (Bioruptor ® Standard, Diagenode, Denville, New Jersey, USA) to an average size of 500 bp. 30 µl of the sheared product was run on a 1% agarose (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, USA) 1x TBE gel to determine size. The remaining sheared volume for each sub-library was concentrated to 30 µl using Mini-elute columns (QIAGEN) and then purified with 1.0X Agencourt AMPure XP beads (Beckman Coulter, Inc., Brea, California, USA) to remove short fragments and adapter dimers. To polish the ends of the sheared DNA, a Quick Blunting Kit (NEB) was used in a 25 µl reaction volume containing 19 µl of sheared DNA, 2.5 µl 10x Blunting Buffer, 2.5 µl dNTP Mix (1mM) and 1.0 µl Blunt Enzyme Mix. The reactions were incubated at RT for 30 min and the product was purified with 1.0X AMPure XP beads. To add 3’ adenine overhangs to the DNA, 19 µl of the previous purified reaction was incubated at 37°C for 30 min in a 25 µl reaction containing 3 µl (10 U) Klenow Fragment (3’-5’ exo, NEB), 2.5 µl NEB Buffer 2 and 0.5 µl dATP (10 mM, Fermentas, Thermo Fisher Scientific, Inc., Waltham, Massachusetts, USA). The reaction was cooled at RT for 30 min and purified with 1.0X AMPure XP beads. The modified P2 Illumina© adapter (all rights reserved 2006 Illumina, Inc., San Diego, California, USA; see Etter 2010) was added in 25 µl ligation reaction, including 19 µl of DNA, 1.0 µl of P2 adapter, 2.5 µl of 10x NEB Buffer 2, 0.25 µl rATP (100mM, Promega), 0.5 µl (1000 U) T4 DNA
ligase (2 million U/ml, NEB), and 1.75 µl H₂O. The reaction was incubated at RT for 30 min and purified with 1.0X AMPure XP beads. Purified sub-libraries were eluted in 51 µl and quantified using the Qubit™ High-Sensitivity dsDNA Assay. 120 ng (~ 10 ng/individual) of each sub-library was used as template in a 100 µl PCR amplification with 50 µl Phusion Hot start Flex 2X Master Mix (NEB) 2X Master Mix and 4 µl modified Illumina© amplification primer mix (10 µM, long-P1-forward primer: 5′-
AATGATACGGCGACCACGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATC*T3′, short-P2-reverse primer: 5′-CAAGCAGAAGACGGCATACG*A3′). PCR conditions included an initial denaturing step at 98°C for 3 min, then 14 cycles of 40 sec at 98°C, 15 sec at 65°C, and 30 sec at 72°C followed by a final extension at 72°C for 5 min.10 µl of the sheared product was run on a 1% agarose 1x TBE gel to check the final sub-library size. The remaining amplified product was purified with AMPure XP beads, including a 0.5X size-exclusion step prior to 1.0X purification to equalize the final size of the 4 sub-libraries. Sub-libraries were quantified using Qubit™ 2.0 High-Sensitivity dsDNA Assay, pooled in equimolar concentrations for a final concentration of 10 nM and sequenced in a single lane of the HiSeq™ 2000 following a rapid sequencing Illumina protocol for 150bp single-end reads at the University of Oregon Genomics Core Facility.

**Processing**

- **Pipeline**- We processed the sequence data, clustered the reads into RAD loci, identified the RAD loci containing SNPs and genotyped all samples using several modules from the STACKS software package version 1.11 (Catchen et al. 2013b; Catchen et al. 2011). First, we sorted the read pairs by barcode, filtered the read quality and removed any read that did not contain both a correct barcode and the remaining six bases of the SbfI restriction site sequence. The 24 red colobus samples were then analyzed as follows. We identified RAD loci and called SNPs in each individual, created a catalog for all loci across individuals and matched each sample against the previously created catalog using denovomap.pl. We used a minimum depth of coverage (-m) of 3 to create stacks, a maximum
nucleotide distance between stacks (-M) of 4 when identifying the loci, and a maximum number of mismatches between sample tags (-n) of 2 when generating the catalog. These parameters were set empirically following Catchen et al. (2013b). We enabled the Deleveraging and Removal algorithms to filter out highly repetitive (likely paralogous) loci. These analyses were carried out on the University of Oregon Applied Computational Instrument for Scientific Synthesis (ACISS) server. Then, we used load_radtags.pl and index_radtags.pl to locally populate and index a MYSQL database of loci. Using the scripts export_sql.pl and populations.pl, we selected the polymorphic loci that were present in 22 individuals (90%). In addition, using the script populations.pl we calculated several genetic diversity parameters (expected and observed heterozygosity, nucleotide diversity and inbreeding coefficient) based on one SNP for each RAD locus for each of the two social groups and the whole population.

• **Runs:** Red colobus sequence files (fastq) for the samples have been made available through the Sequence Read Archive (SRA) repository at NCBI (SRP050135)

**Results**

• Information on the RAD loci identified was submitted to Dryad (http://dx.doi.org/10.5061/dryad.404fh), including information of each locus present in at least 22 individuals and consensus genotypes for each individual.

• Table 1 shows the number of reads obtained in the run and after the quality filtering steps.

• Table 2 shows number of reads obtained after quality filtering per individual, in addition to specific data for each individual, such as sex, social group and date of sampling.

• Table 3 shows several genetic diversity parameters for both social groups and overall. Where Ho is observed heterozygosity, He is expected heterozygosity, π is nucleotide diversity, and Fis is the inbreeding coefficient of an individual relative to the subpopulation.

• Quality scoring system: phred + 33
Acknowledgments
This research was supported by NIH grant TW009237 as part of the joint NIH-NSF Ecology of Infectious Disease program and the UK Economic and Social Research Council and a University of Oregon Research, Innovation and Graduate Education Seeding Success Faculty Research Award. Use of trade names is for identification only and does not imply endorsement by the U.S. Department of Health and Human Services, the Public Health Service, or the Centers for Disease Control and Prevention. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

References


Table 1. Number of reads obtained in the run and after each filtering step.

<table>
<thead>
<tr>
<th></th>
<th>Number of Reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library</td>
<td>154,469,192</td>
</tr>
<tr>
<td>Unidentified barcodes</td>
<td>11,399,135</td>
</tr>
<tr>
<td>Other primate species(^1)</td>
<td>72,296,200</td>
</tr>
<tr>
<td>Red colobus</td>
<td>70,773,857</td>
</tr>
<tr>
<td>Retained after quality filtering</td>
<td>58,814,906</td>
</tr>
<tr>
<td>Mean per individual</td>
<td>2,948,911</td>
</tr>
</tbody>
</table>

\(^1\) Reads corresponding to the other primate species that were analyzed in the same library than the Red Colobus individuals.
Table 2. Number of reads obtained for each of the red colobus individuals sequenced using RAD-Seq.

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>Social Group</th>
<th>Date</th>
<th>Number of reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>R11</td>
<td>Female</td>
<td>Small Camp</td>
<td>2006</td>
<td>2444654</td>
</tr>
<tr>
<td>R12</td>
<td>Male</td>
<td>Large Mikana</td>
<td>2006</td>
<td>2623855</td>
</tr>
<tr>
<td>R13</td>
<td>Female</td>
<td>Large Mikana</td>
<td>2006</td>
<td>2345512</td>
</tr>
<tr>
<td>R21</td>
<td>Male</td>
<td>Large Mikana</td>
<td>2006</td>
<td>1272056</td>
</tr>
<tr>
<td>R30</td>
<td>Female</td>
<td>Small Camp</td>
<td>2006</td>
<td>4180405</td>
</tr>
<tr>
<td>R31</td>
<td>Female</td>
<td>Small Camp</td>
<td>2006</td>
<td>2219072</td>
</tr>
<tr>
<td>R33</td>
<td>Female</td>
<td>Small Camp</td>
<td>2006</td>
<td>3328019</td>
</tr>
<tr>
<td>R34</td>
<td>Female</td>
<td>Small Camp</td>
<td>2006</td>
<td>2175606</td>
</tr>
<tr>
<td>R40</td>
<td>Male</td>
<td>Small Camp</td>
<td>2006</td>
<td>3696437</td>
</tr>
<tr>
<td>R41</td>
<td>Female</td>
<td>Large Mikana</td>
<td>2006</td>
<td>2277670</td>
</tr>
<tr>
<td>R49</td>
<td>Male</td>
<td>Large Mikana</td>
<td>2006</td>
<td>2753231</td>
</tr>
<tr>
<td>R52</td>
<td>Female</td>
<td>Large Mikana</td>
<td>2006</td>
<td>2184823</td>
</tr>
<tr>
<td>R54</td>
<td>Female</td>
<td>Small Camp</td>
<td>2006</td>
<td>2281773</td>
</tr>
<tr>
<td>R56</td>
<td>Female</td>
<td>Small Camp</td>
<td>2006</td>
<td>4372857</td>
</tr>
<tr>
<td>R6</td>
<td>Male</td>
<td>Large Mikana</td>
<td>2006</td>
<td>2351825</td>
</tr>
<tr>
<td>R65</td>
<td>Female</td>
<td>Large Mikana</td>
<td>2006</td>
<td>2120931</td>
</tr>
<tr>
<td>R67</td>
<td>Female</td>
<td>Large Mikana</td>
<td>2006</td>
<td>1948210</td>
</tr>
<tr>
<td>R70</td>
<td>Female</td>
<td>Large Mikana</td>
<td>2006</td>
<td>1466084</td>
</tr>
<tr>
<td>R72</td>
<td>Female</td>
<td>Large Mikana</td>
<td>2006</td>
<td>3080133</td>
</tr>
<tr>
<td>R73</td>
<td>Female</td>
<td>Small Camp</td>
<td>2006</td>
<td>1985157</td>
</tr>
<tr>
<td>R996</td>
<td>Male</td>
<td>Large Mikana</td>
<td>2006</td>
<td>2623420</td>
</tr>
<tr>
<td>R999</td>
<td>Male</td>
<td>Small Camp</td>
<td>2006</td>
<td>1717138</td>
</tr>
<tr>
<td>RC54</td>
<td>Female</td>
<td>Small Camp</td>
<td>2010</td>
<td>1899120</td>
</tr>
<tr>
<td>RC60</td>
<td>Male</td>
<td>Large Mikana</td>
<td>2010</td>
<td>1466918</td>
</tr>
</tbody>
</table>
Table 3. Mean and standard deviation estimates for heterozygosity, nucleotide diversity, and inbreeding coefficient.

<table>
<thead>
<tr>
<th></th>
<th>Ho</th>
<th>He</th>
<th>π</th>
<th>Fis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large Mikana</td>
<td>0.242</td>
<td>0.248</td>
<td>0.259</td>
<td>0.045</td>
</tr>
<tr>
<td>(SD)</td>
<td>0.184</td>
<td>0.165</td>
<td>0.172</td>
<td>0.255</td>
</tr>
<tr>
<td>Small Camp</td>
<td>0.252</td>
<td>0.249</td>
<td>0.262</td>
<td>0.027</td>
</tr>
<tr>
<td>(SD)</td>
<td>0.192</td>
<td>0.166</td>
<td>0.174</td>
<td>0.253</td>
</tr>
<tr>
<td>Both groups</td>
<td>0.247</td>
<td>0.255</td>
<td>0.261</td>
<td>0.045</td>
</tr>
<tr>
<td>(SD)</td>
<td>0.167</td>
<td>0.157</td>
<td>0.161</td>
<td>0.216</td>
</tr>
</tbody>
</table>

Ho is observed heterozygosity, He is expected heterozygosity, π is nucleotide diversity, and Fis is the inbreeding coefficient of an individual relative to the subpopulation.