AN INEXPENSIVE METHYLATION-BASED ENRICHMENT METHOD ENABLES GENOMIC-SCALE POPULATION-LEVEL GENOTYPING OF ANIMALS FROM THEIR FECES

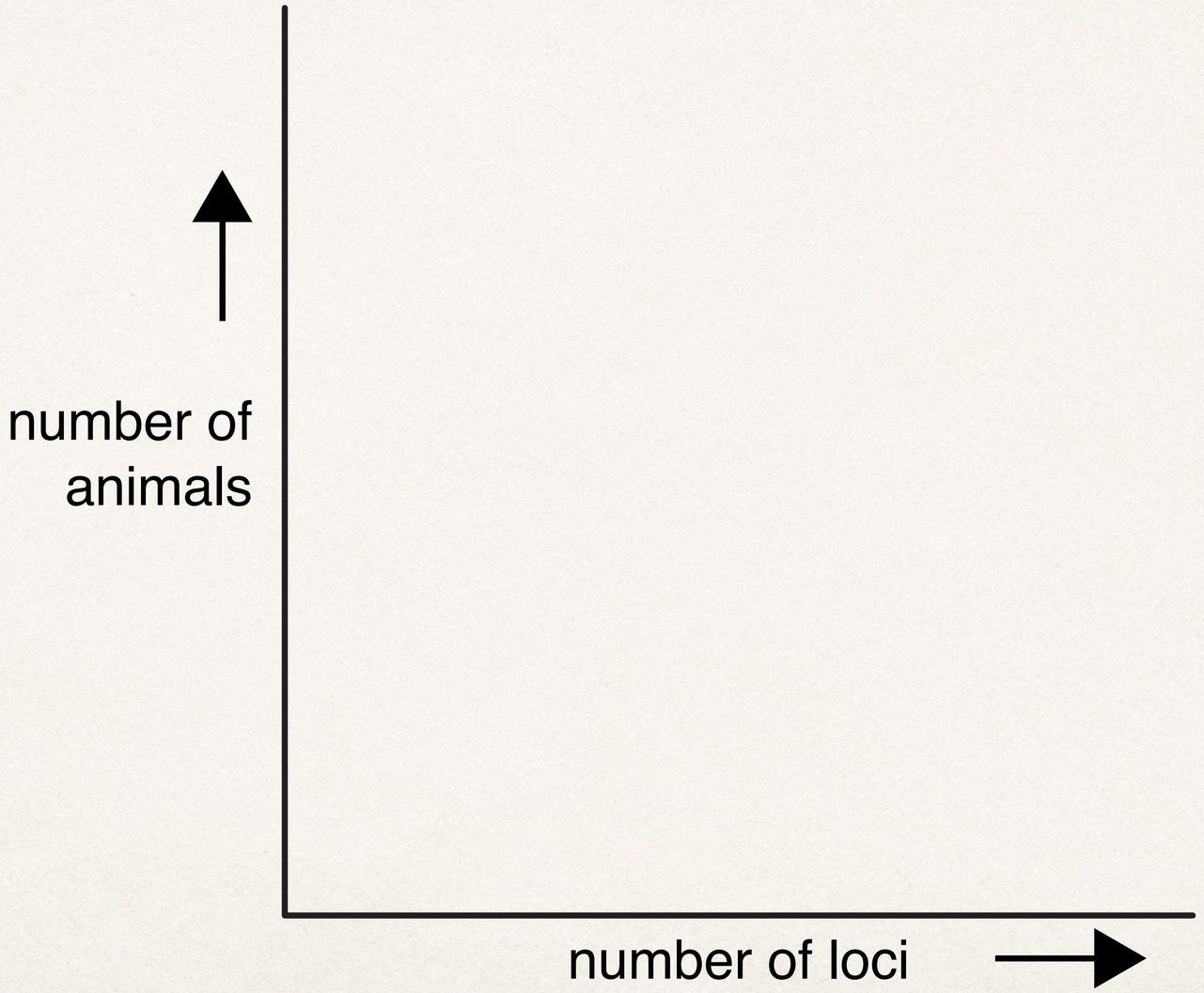
Kenneth L. Chiou and Christina M. Bergey

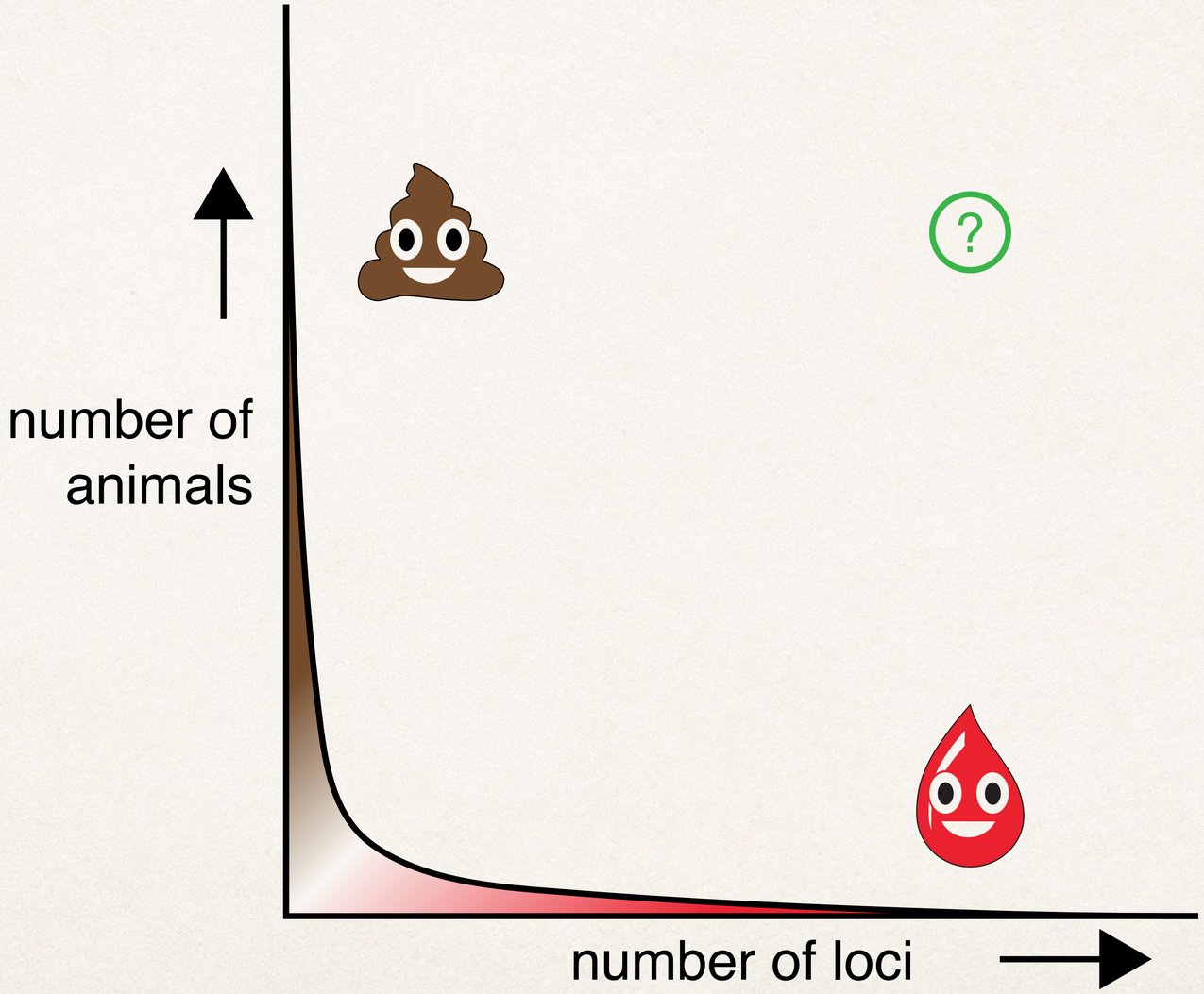
Washington University in St. Louis

NYU

NYCEP

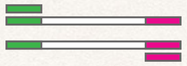
UNIVERSITY OF NOTRE DAME


number of
animals
number of loci


number of animals
?
number of loci

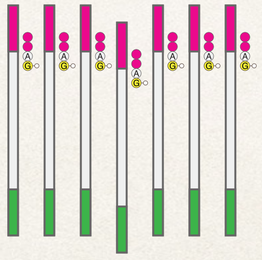
Most of what we know about the genetics of primate populations comes from **first-generation sequencing** of DNA from **feces**

Primate cells are present in their feces



Target DNA

Copy DNA



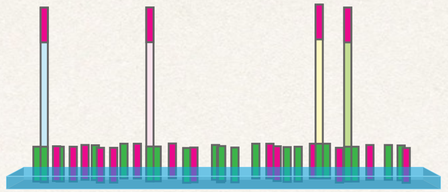
Read DNA

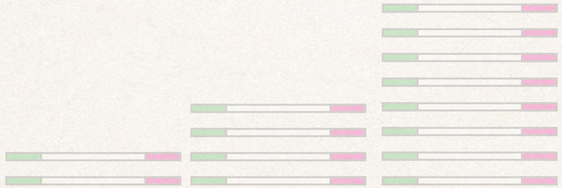
If this screen represents the human genome (3 billion base pairs) . . .

1 thousand base pairs (red)

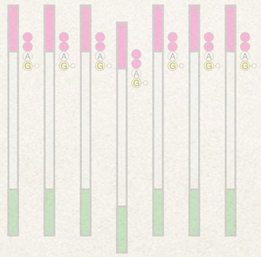
Most of what we are learning about the genomics of primates is coming from **second-generation sequencing** of DNA

Isolate DNA
in parallel

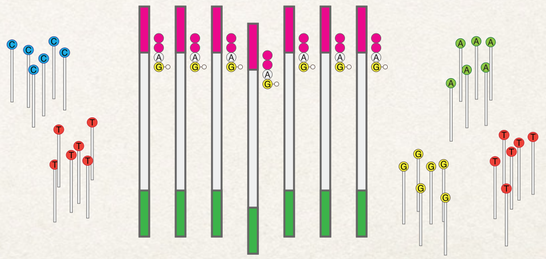


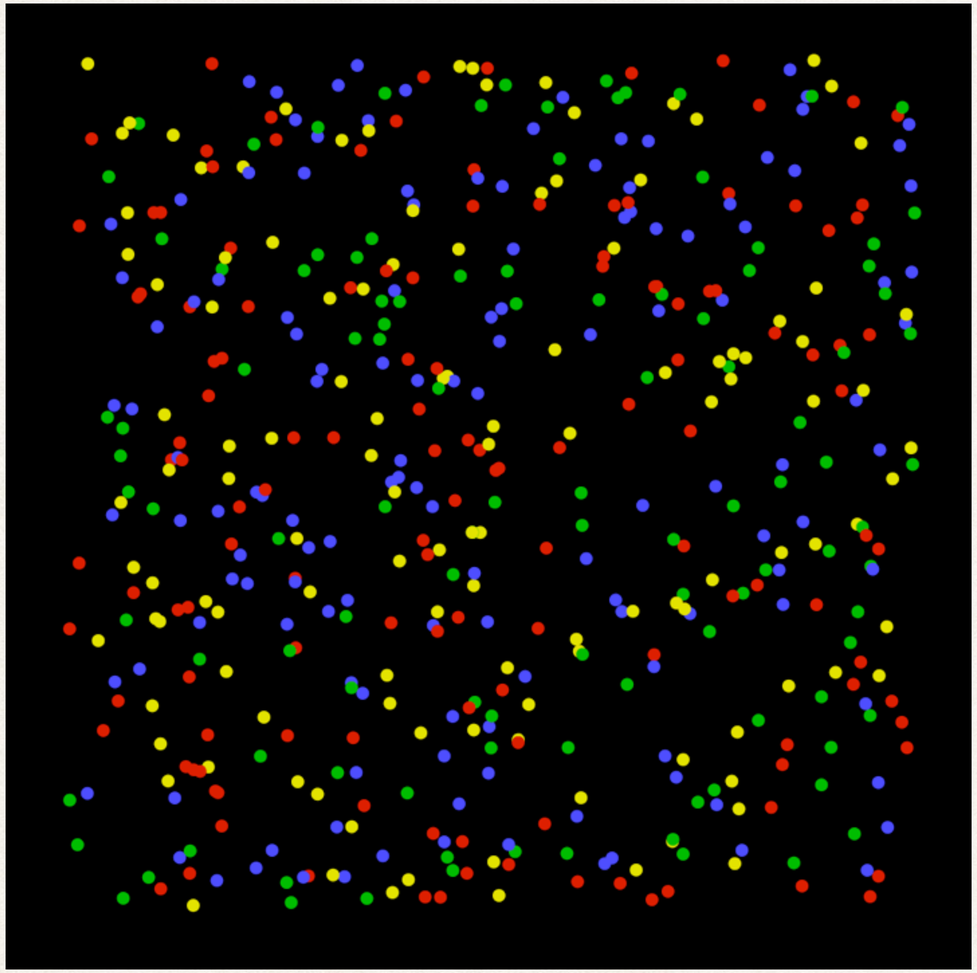


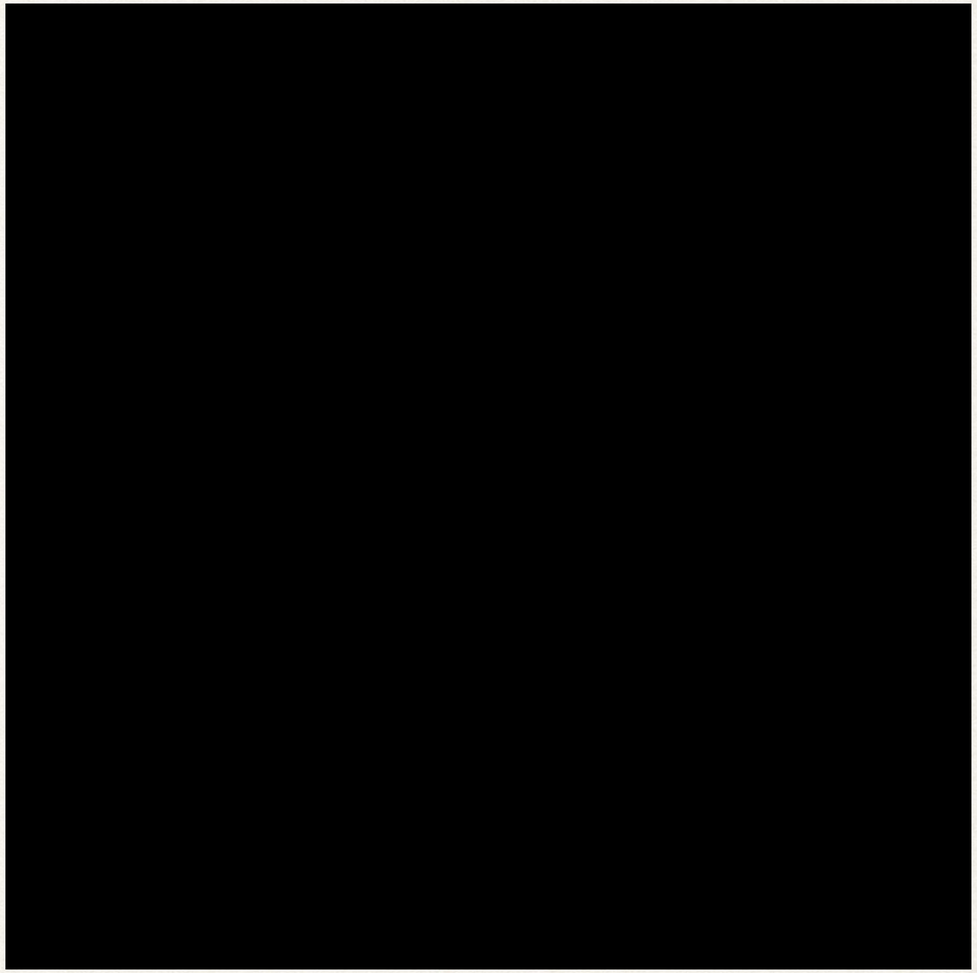
Copy DNA
in parallel



Read DNA
in parallel

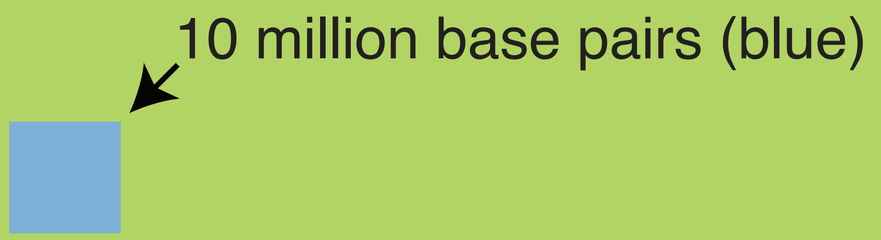






Is the whole genome too much data?

If this screen represents the human genome (3 billion base pairs) . . .



Given the same budget . . .

1 sample

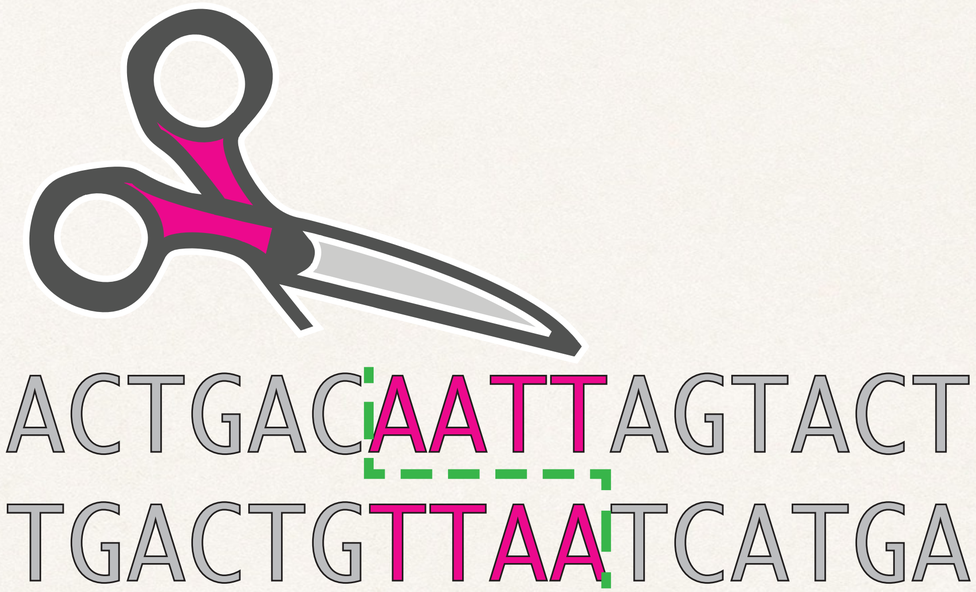
3 billion base pairs

300 samples

10 million base pairs each

Reduced representation genome

double-digest RAD sequencing



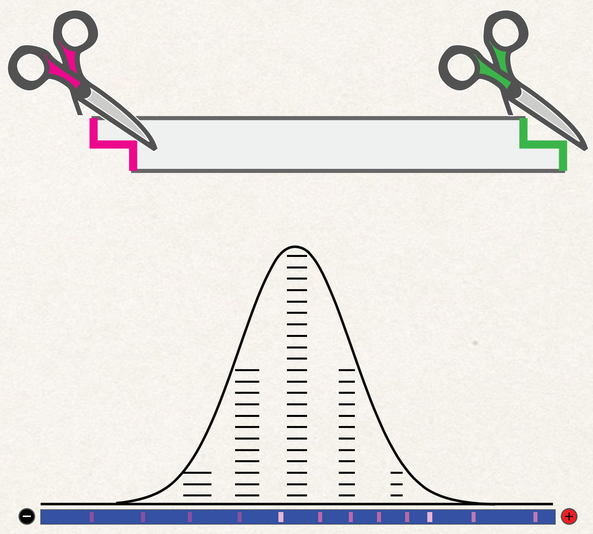
Restriction enzyme

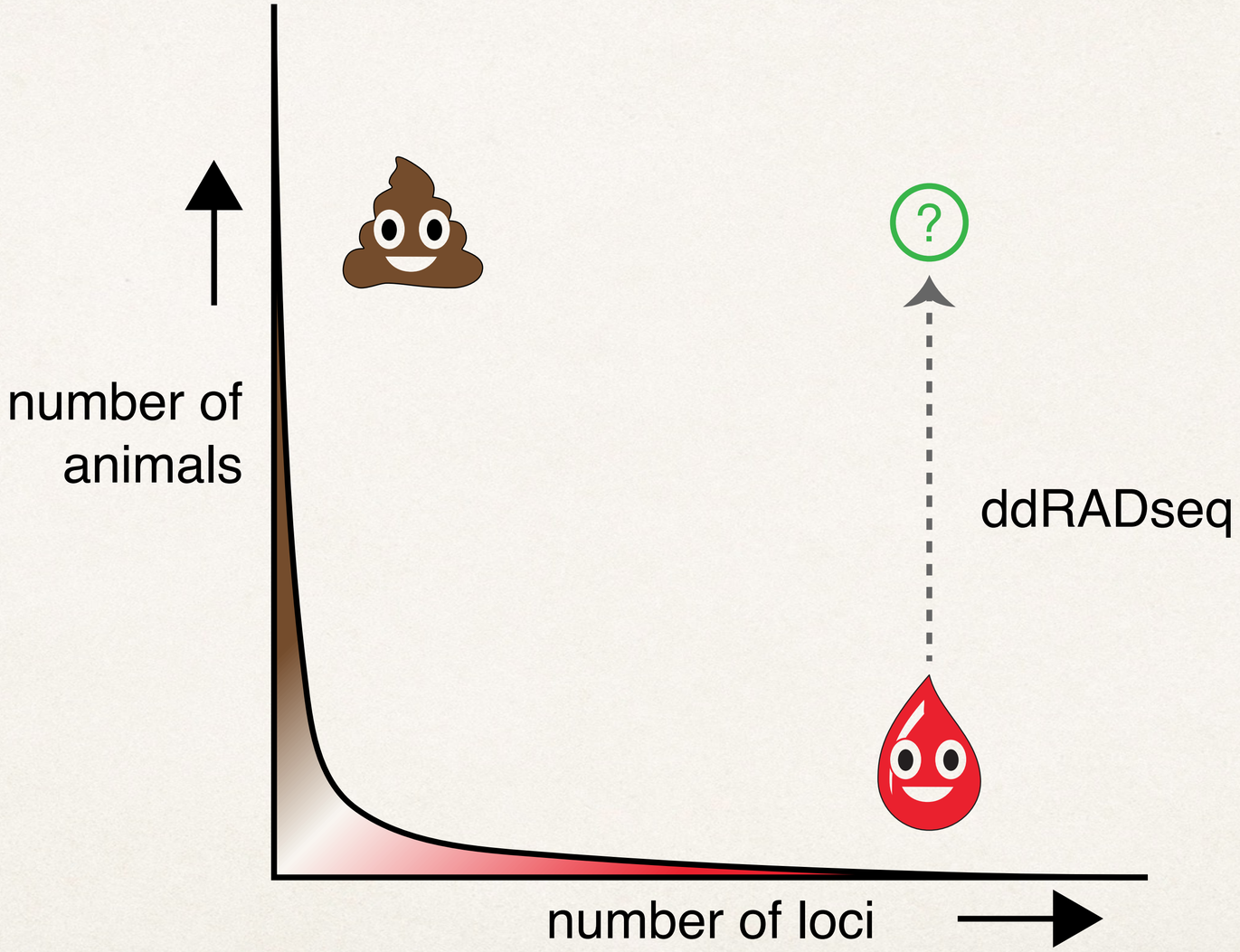
ddRADseq

Reduces genome to a subset of markers based on two criteria:

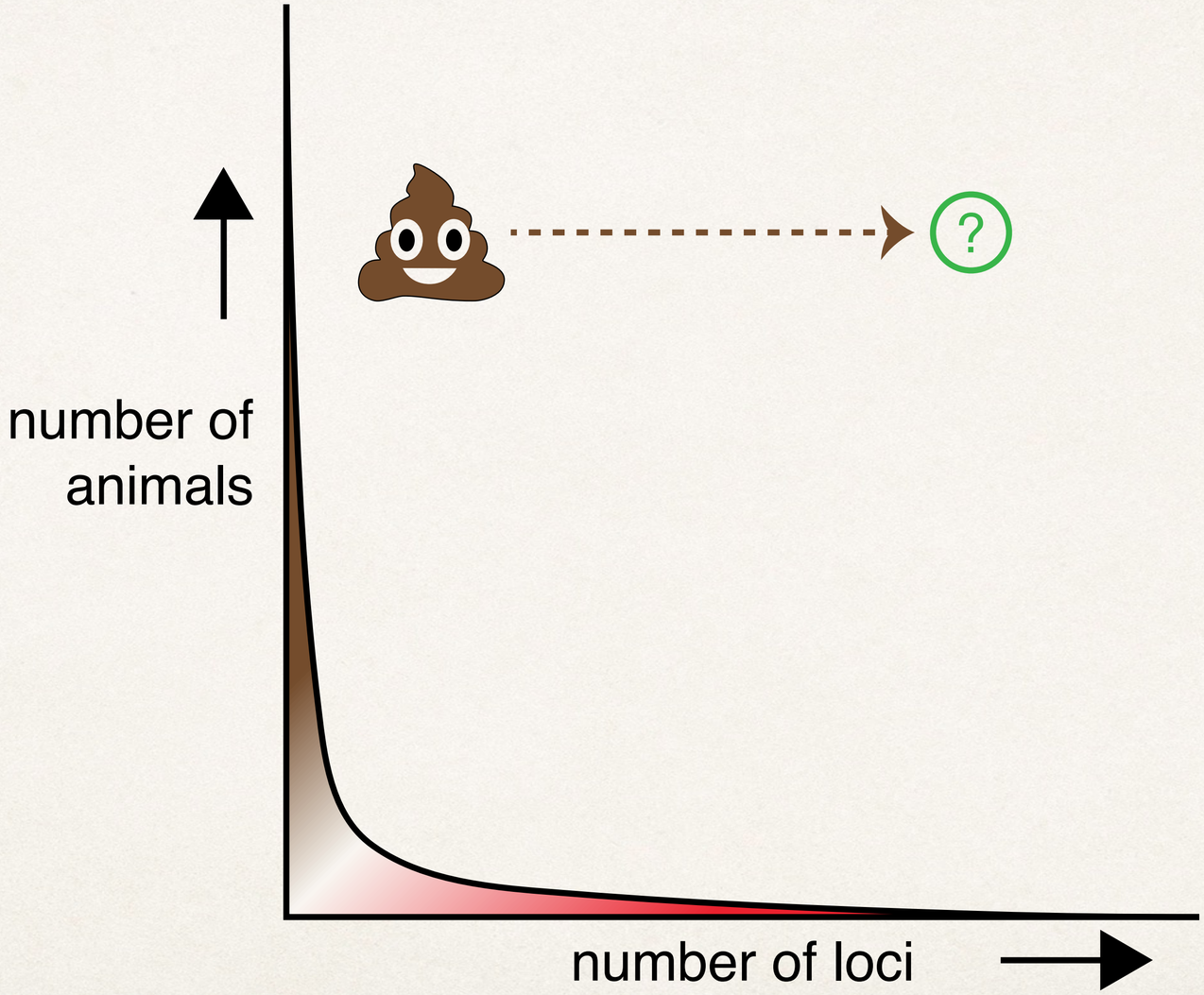
- flanked by both enzyme cut sites
- specified length

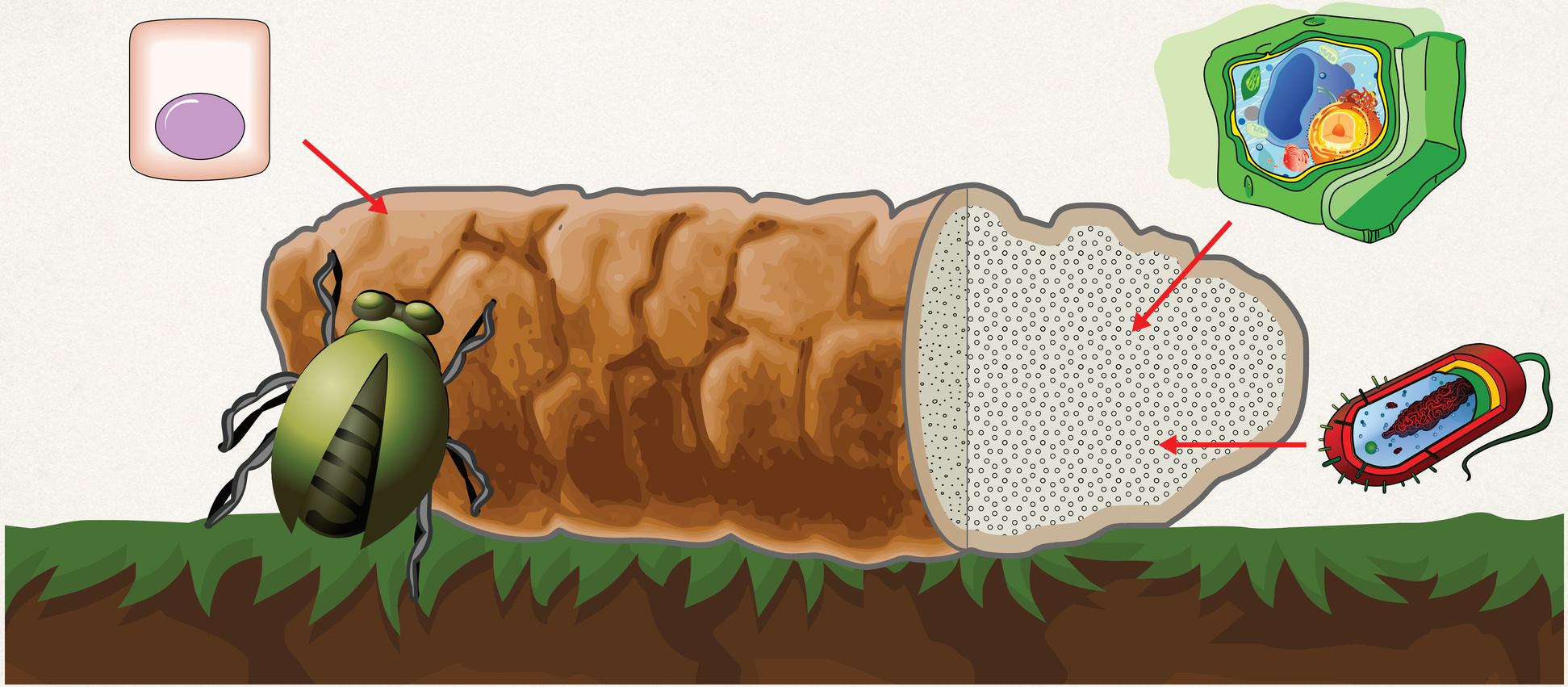
Marker set can be tuned through choice of enzymes and fragment length



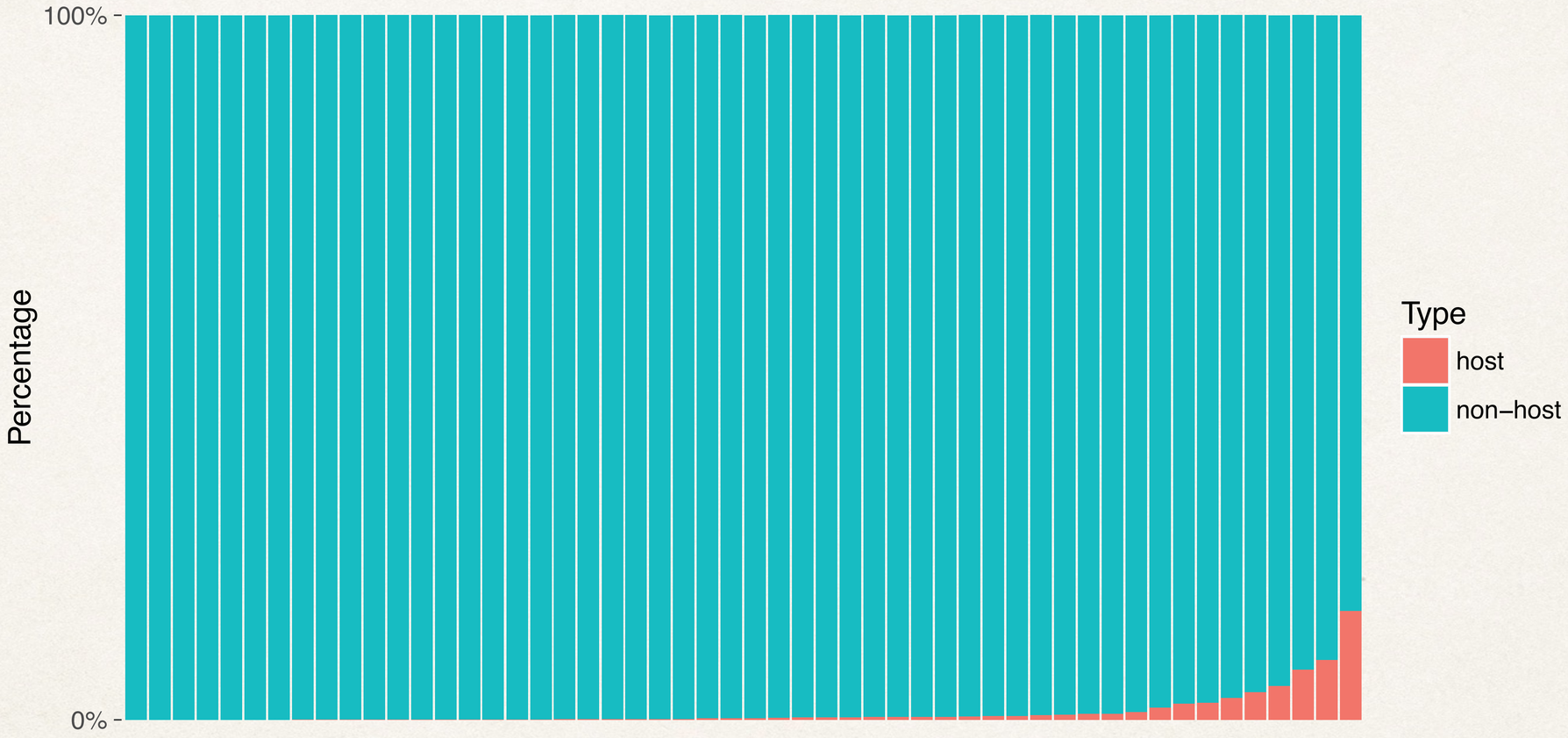

number of animals
number of loci
?
ddRADseq

Kafue National Park, Zambia


number of animals
?
number of loci

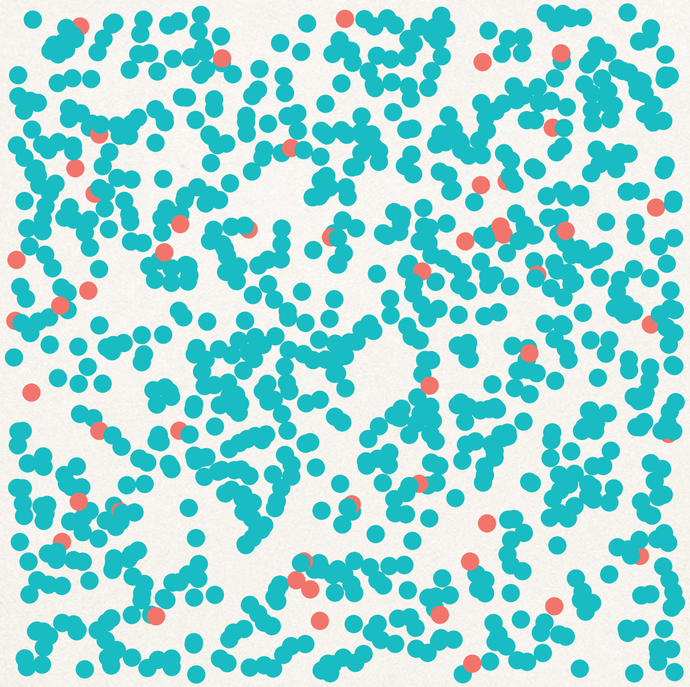


Exogenous sources of DNA

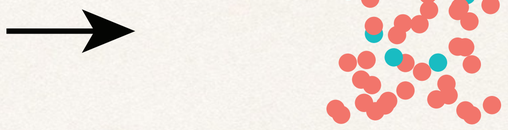


Percentage of host DNA in fecal DNA

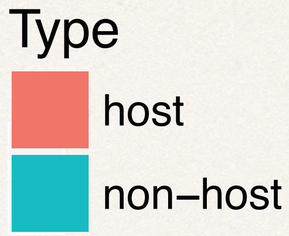
Genome-scale enrichment



fecal DNA



enriched host DNA



Approaches

- PCR (not genome-scale)
- Targeted sequence capture (Perry *et al.* 2010; Snyder-Mackler *et al.* 2016)
- CpG methylation capture (*this study*)

Approaches

- PCR (not genome-scale)
- Targeted sequence capture (Perry *et al.* 2010; Snyder-Mackler *et al.* 2016)
- CpG methylation capture (*this study*)

MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 5332–5344 doi: 10.1111/j.1365-294X.2010.04888.x

FROM THE COVER

Genomic-scale capture and sequencing of endogenous DNA from feces

GEORGE H. PERRY[1], JOHN C. MARIONI[1], PÁLL MELSTED and YOAV GILAD

Department of Human Genetics, University of Chicago, 920 E. 58th St., Chicago, IL 60637, USA

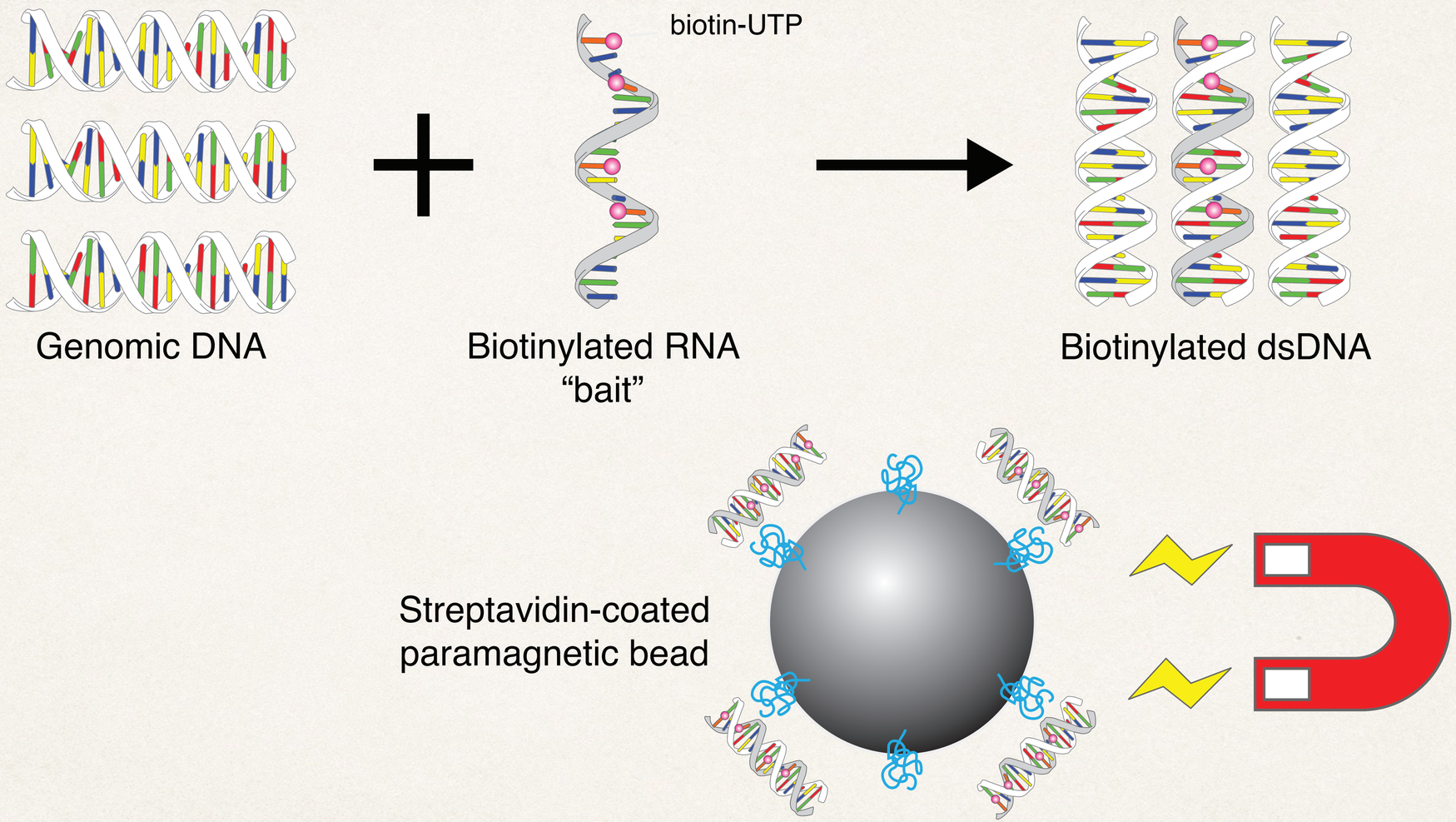
HIGHLIGHTED ARTICLE

GENETICS | INVESTIGATION

Efficient Genome-Wide Sequencing and Low-Coverage Pedigree Analysis from Noninvasively Collected Samples

Noah Snyder-Mackler,* William H. Majoros,† Michael L. Yuan,* Amanda O. Shaver,* Jacob B. Gordon,‡ Gisela H. Kopp,§,,†† Stephen A. Schlebusch,‡‡ Jeffrey D. Wall,§§ Susan C. Alberts,*,‡,*** Sayan Mukherjee,†††,‡‡‡,§§§ Xiang Zhou,****,††††,1 and Jenny Tung*,‡,***,‡‡‡‡,1**

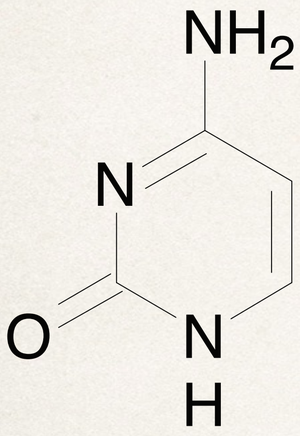
*Department of Evolutionary Anthropology, †Graduate Program in Computational Biology and Bioinformatics, ‡Department of Biology, †††Department of Statistical Science, ‡‡‡Department of Mathematics, §§§Department of Computer Science, and ‡‡‡‡Duke University Population Research Institute, Duke University, Durham, North Carolina 27708, §Cognitive Ethology Laboratory, German Primate Center, Leibniz Institute for Primate Research, 37077 Göttingen, Germany, **Department of Biology, University of Konstanz, 78457 Konstanz, Germany, ††Department of Migration and Immuno-Ecology, Max Planck Institute for Ornithology, 82319 Radolfzell, Germany, ‡‡Department of Molecular and Cell Biology, University of Cape Town, 7700 Cape Town, South Africa, §§Institute for Human Genetics, University of California, San Francisco, California 94143, ***Institute of Primate Research, National Museums of Kenya, Nairobi, Kenya, and ****Department of Biostatistics and ††††Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109



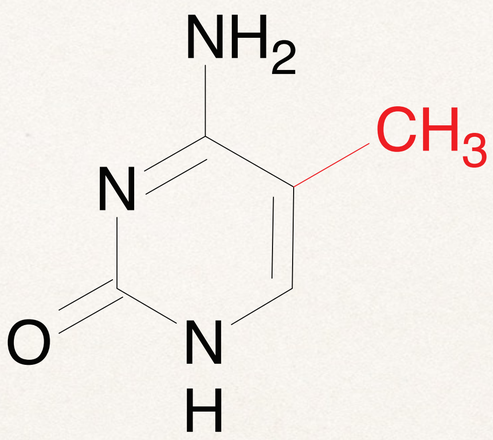
Targeted sequence capture

Approaches

- PCR (not genome-scale)
- Targeted sequence capture (Perry *et al.* 2010; Snyder-Mackler *et al.* 2016)
- CpG methylation capture (*this study*)

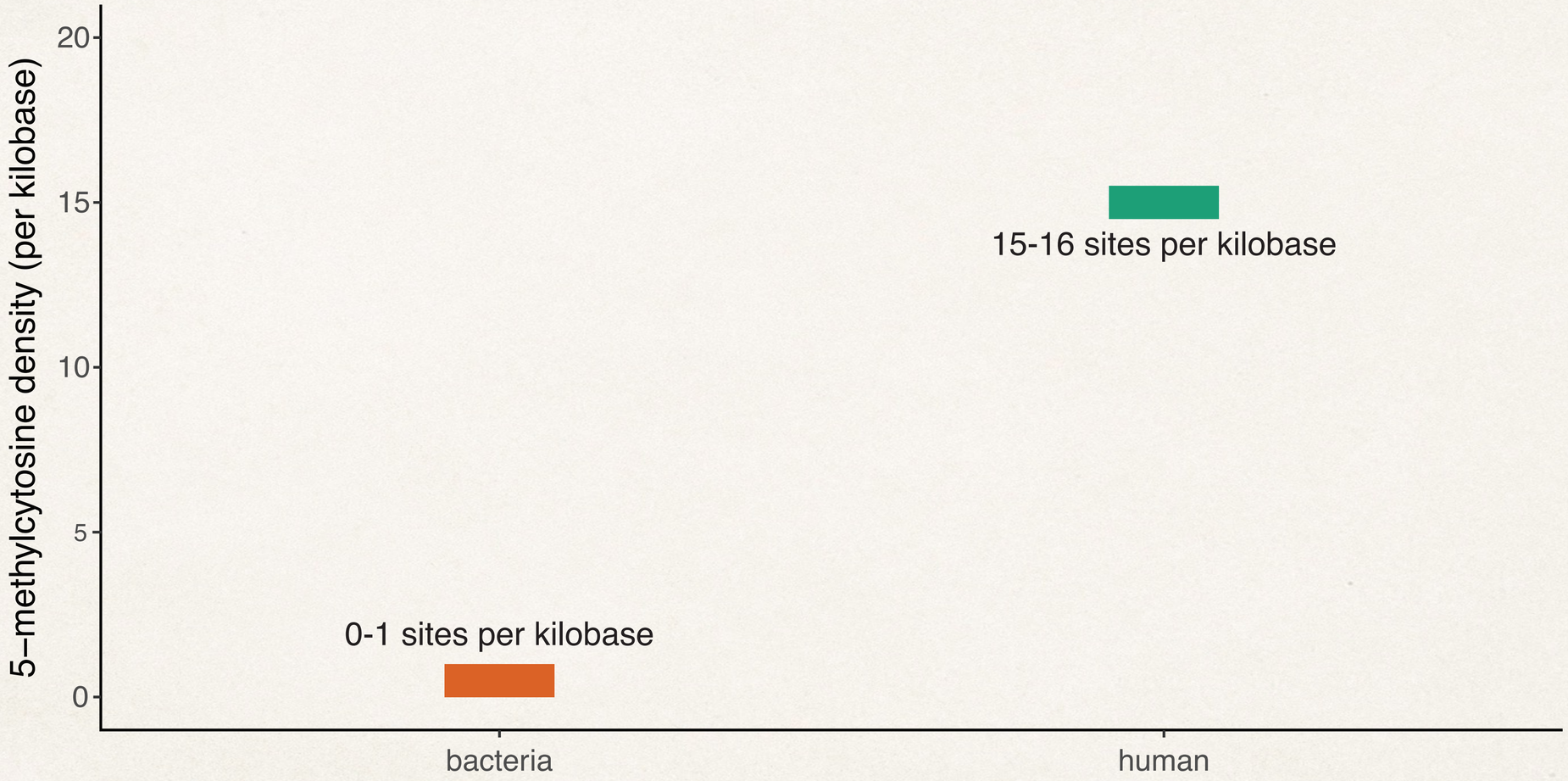


cytosine

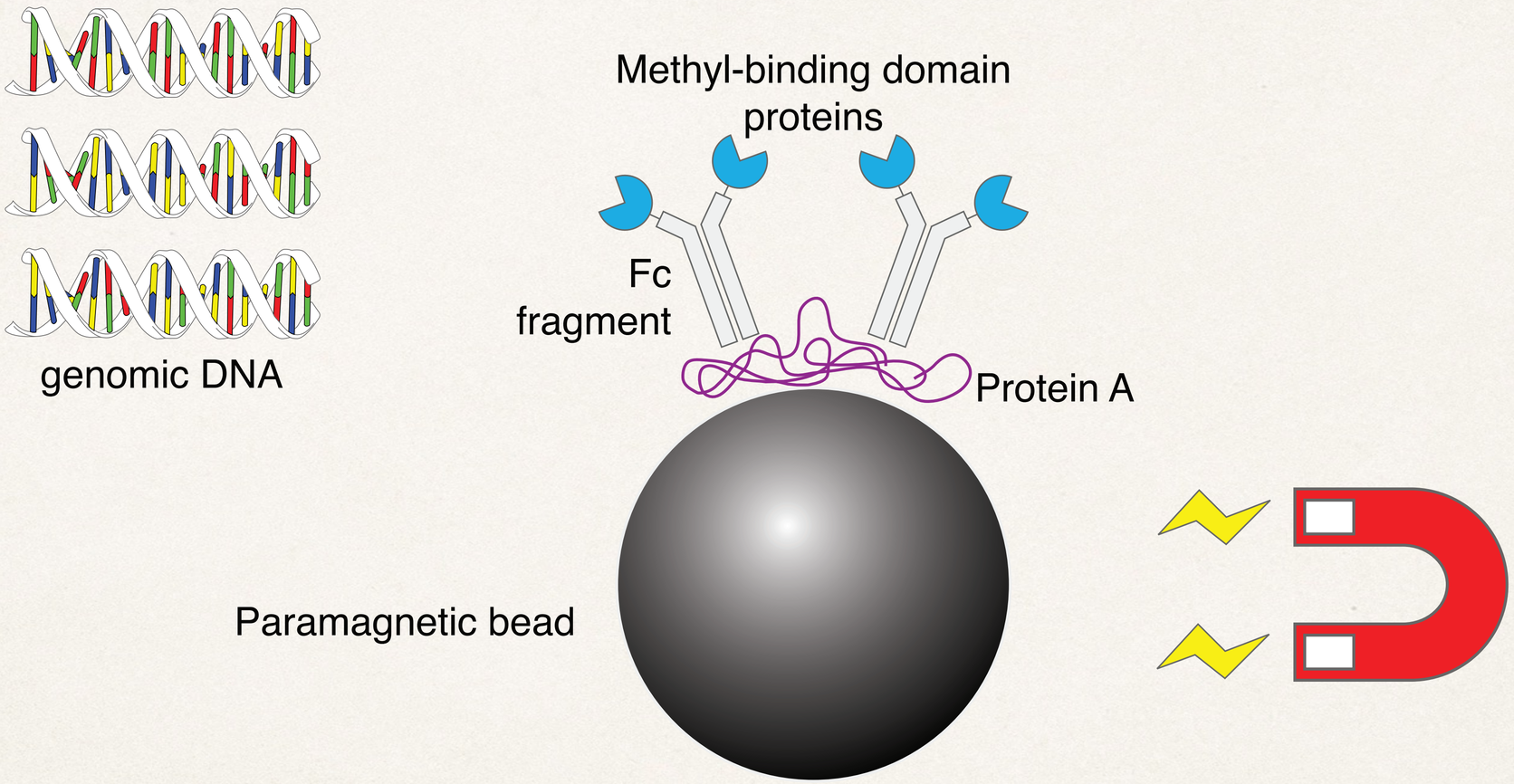


5-methylcytosine

CpG site



Typical CpG methylation densities



CpG methylation capture

OPEN ACCESS Freely available online

PLOS ONE

A Method for Selectively Enriching Microbial DNA from Contaminating Vertebrate Host DNA

George R. Feehery[1], Erbay Yigit[1], Samuel O. Oyola[2], Bradley W. Langhorst[1], Victor T. Schmidt[3,5], Fiona J. Stewart[1], Eileen T. Dimalanta[1], Linda A. Amaral-Zettler[3,4], Theodore Davis[1], Michael A. Quail[2], Sriharsa Pradhan[1]*

1 New England Biolabs Inc., Ipswich, Massachusetts, United States of America, **2** Wellcome Trust Sanger Institute, Cambridge, United Kingdom, **3** The Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, Massachusetts, United States of America, **4** Department of Geological Sciences, Brown University, Providence, Rhode Island, United States of America, **5** Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island, United States of America

Methods

Dataset

- Blood and feces from captive baboons (n=6)
  - *P. anubis*, *P. anubis* x *P. cynocephalus*, *P. anubis* x *P. ursinus*
- Feces from wild Zambian baboons (n=46)
  - *P. kindae*, *P. ursinus griseipes*, *P. kindae* x *P. ursinus griseipes*, *P. kindae* x *P. cynocephalus*
  - 2006 (n=10)
  - 2007 (n=15)
  - 2014 (n=18)
  - 2015 (n=3)

Labwork outline

- Extractions (standard procedures)
- qPCR (standard procedures)
- Enrich host DNA
- ddRADseq (standard procedures)
  - *Sph*I + *Mlu*CI; target = 300bp; Peterson *et al.* 2012
- Sequencing (standard procedures)
  - Illumina MiSeq, HiSeq 2500, HiSeq 4000

Methyl-CpG enrichment

- NEBNext Microbiome DNA Enrichment Kit (New England Biolabs), following published protocol (Feehery *et al.* 2011)
- Enrichment success evaluated by qPCR or Illumina sequencing

Protocol tuning

After two rounds of pilot sequencing

- Simulated fecal DNA by combining baboon gDNA with *E. coli* DNA in controlled proportions
- Performed 120 enrichment trials to evaluate protocol modifications, with success measured by qPCR

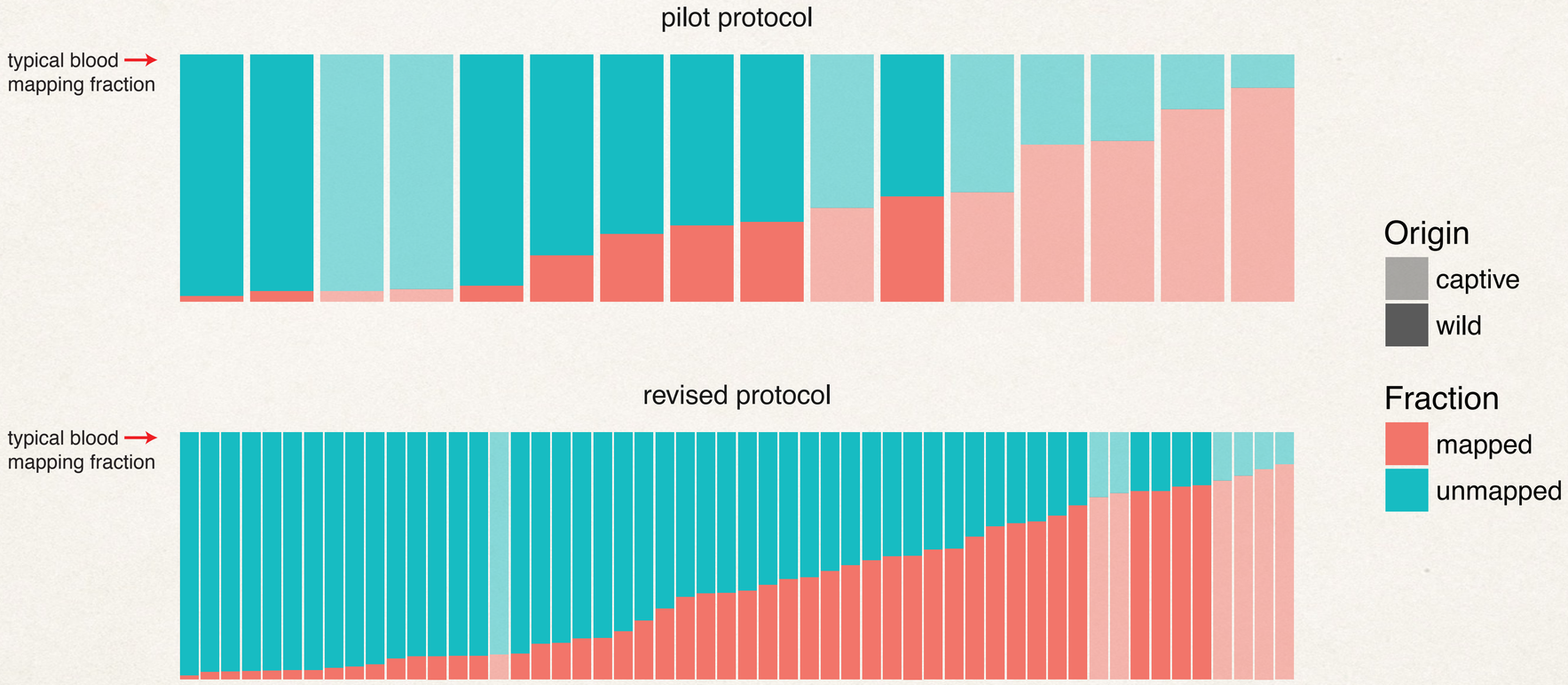
Bioinformatics

- Demultiplex samples (sabre) and trim adapters (ea-utils)
- Map reads to baboon genome (papAnu2) with BWA-MEM
- Identify variants (GATK UnifiedGenotyper)
- Filter SNPs (GATK VariantFiltration)
  - QD < 2.0, MQ < 40.0, FS > 60.0, HaplotypeScore > 13.0, MQRankSum < -12.5, ReadPosRankSum < -8.0; indels excluded
- Calculate statistics (VCFtools; PLINK 1.07)

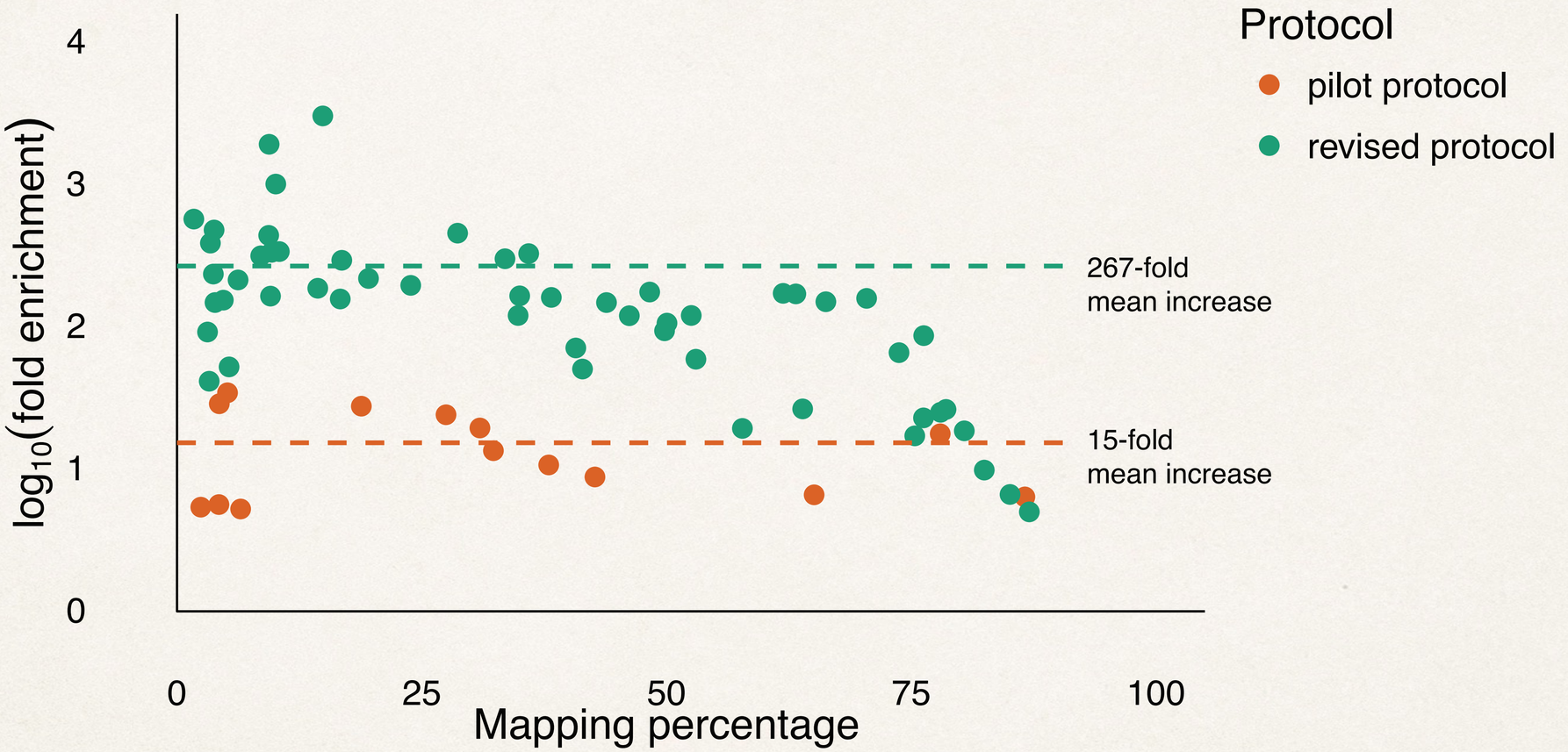
Results

Results

Mapping success



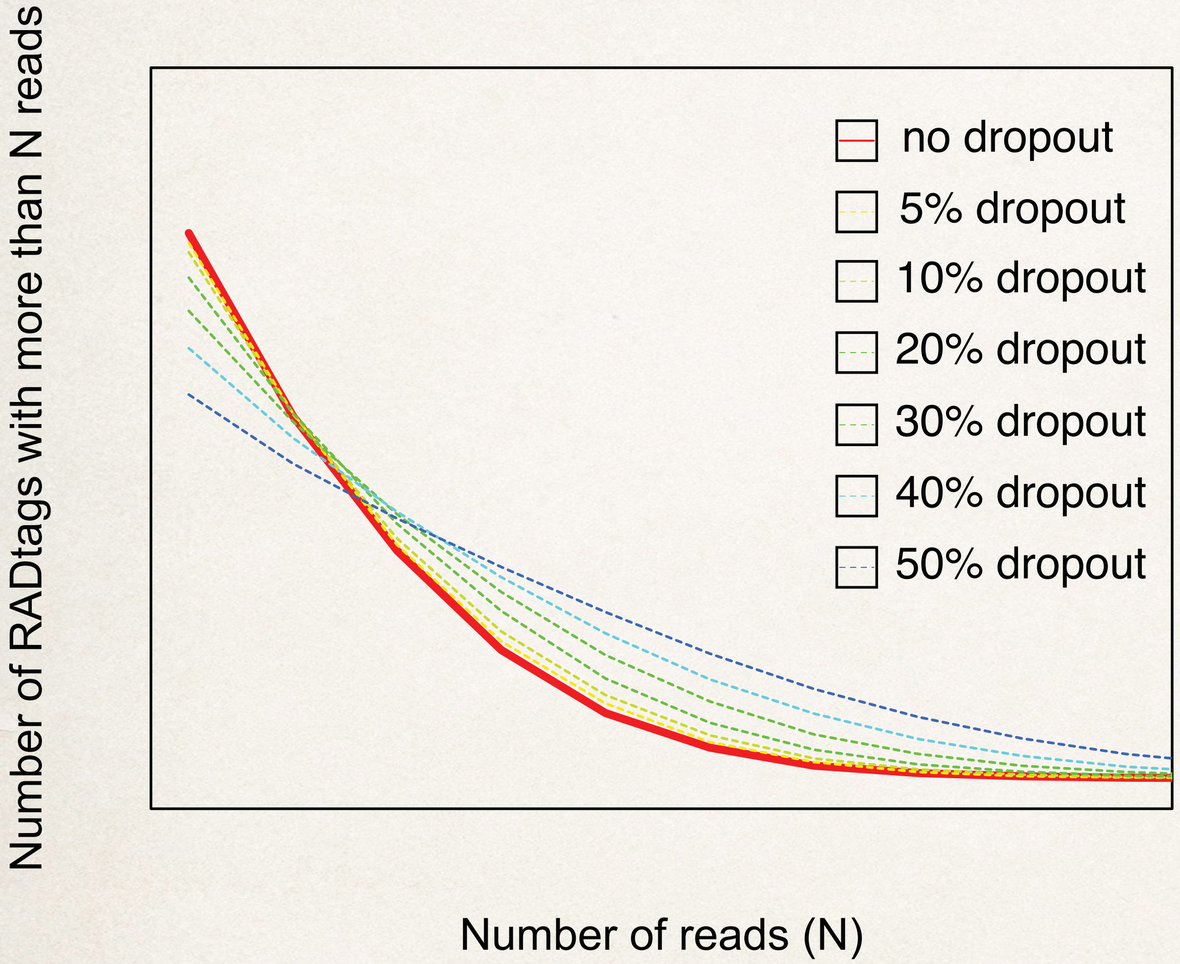
Fraction of reads matching baboon genome

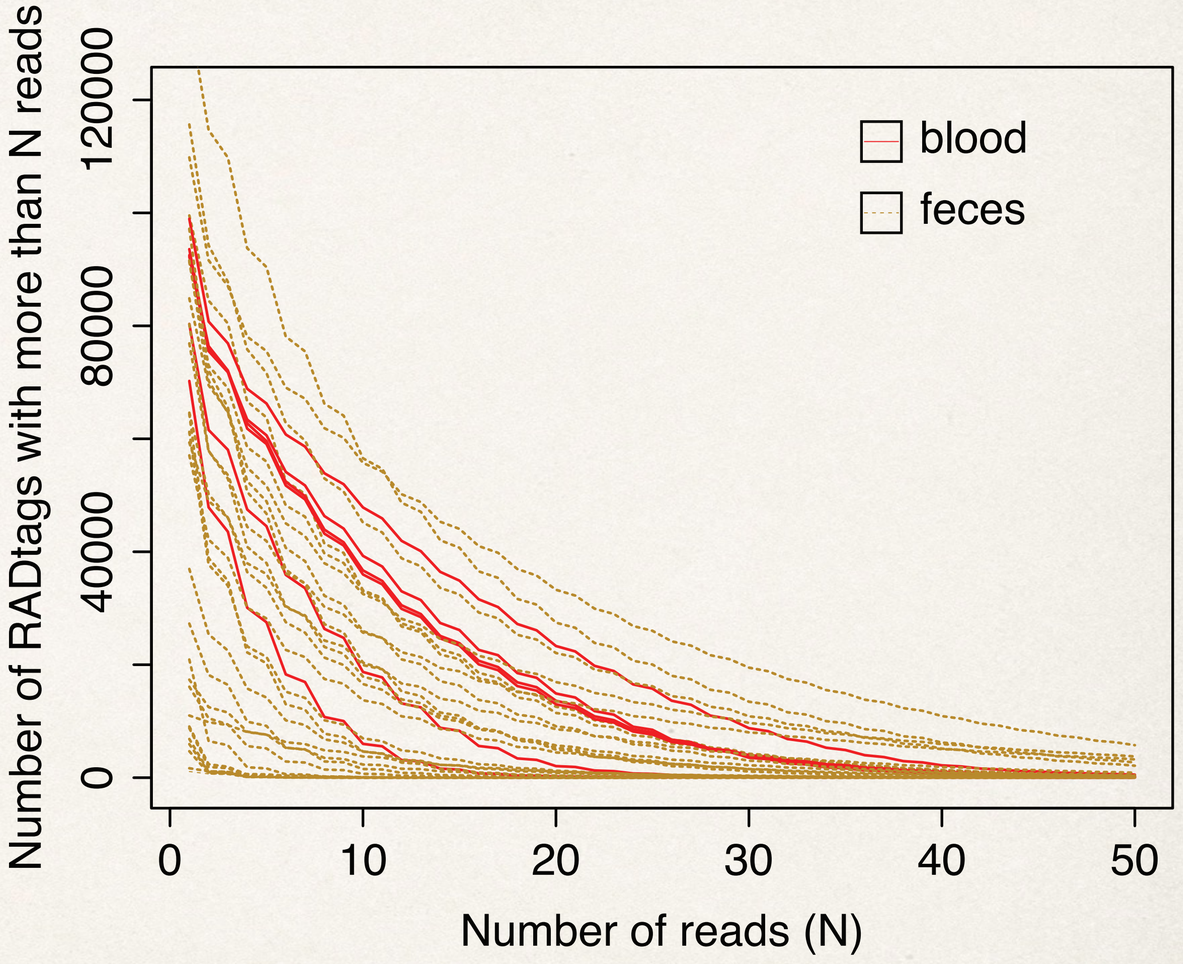


Comparison of *magnitude* of enrichment

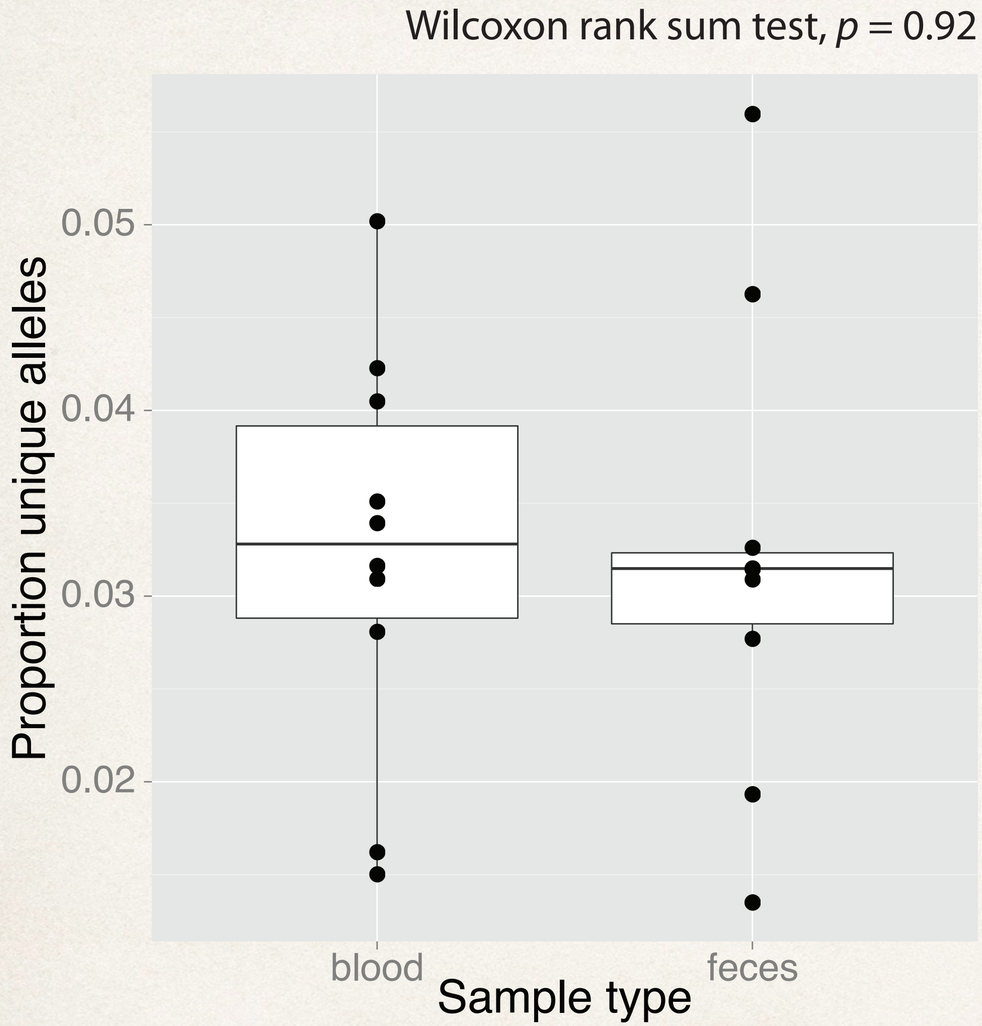
Results

Potential capture biases

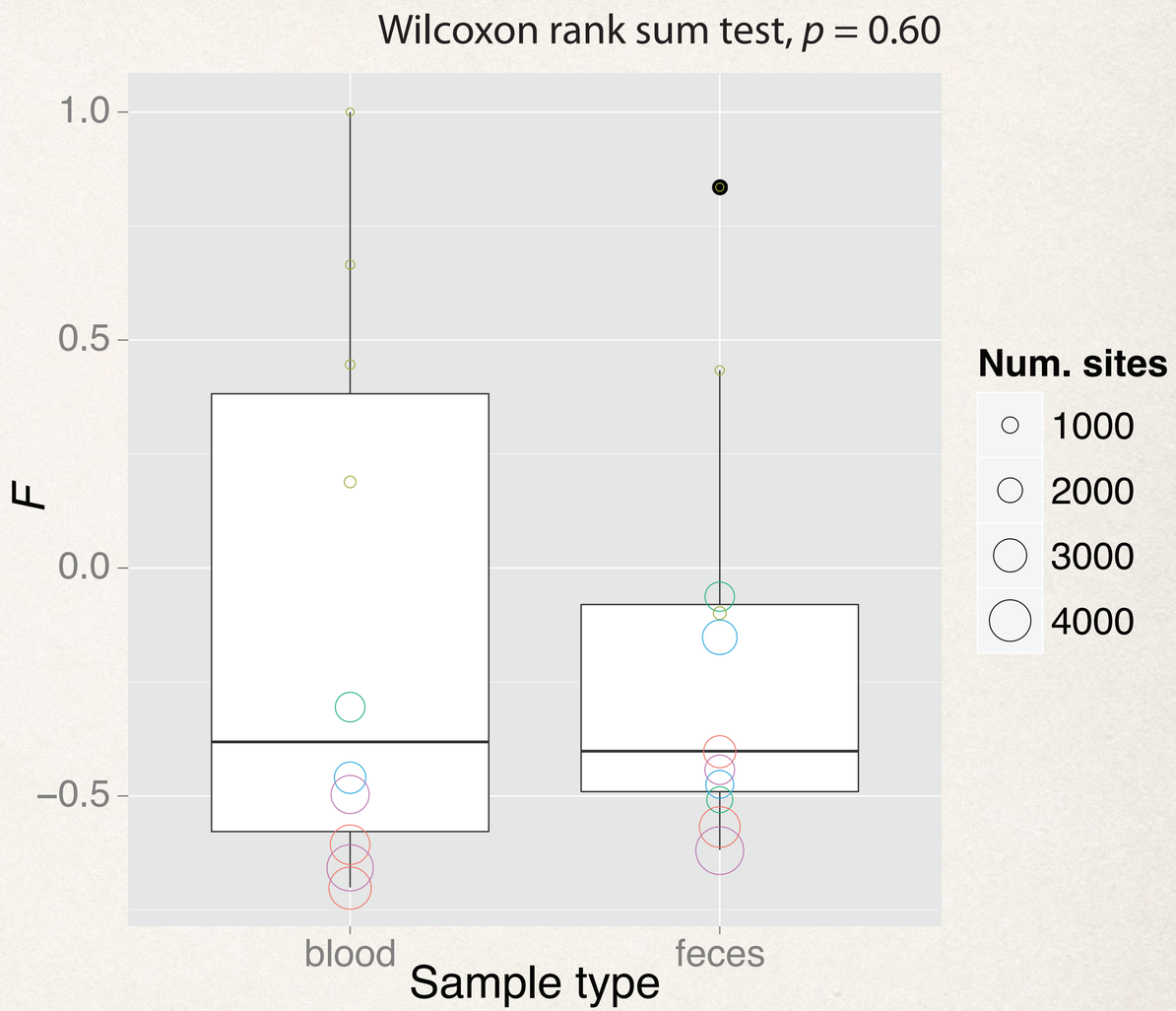




Cumulative coverage plots



Proportion of unique alleles



Inbreeding coefficient

Discussion

	Snyder-Mackler *et al.* 2016	This study
Baits	transcribed RNA	methyl-binding domain protein
Genome mapping rate (optimized protocols only; PCR duplicates included)	56.73% (from 2.80%)	32.49% (from 0.51%)
Enrichment cost (per sample)	$29	$0.70
IPS/ASP 2016 talk	**Tomorrow at 6:15pm (room 327)**	Right now (right here)

Pros

- Fast and cheap (< 1 day of work)
- Homogeneous protein bait
  - Bait preparation and hybridization are highly replicable
  - No issues with cross-taxon sequences
- Long strands are captured without modification
  - Readily compatible with downstream library preparation methods, including ddRADseq

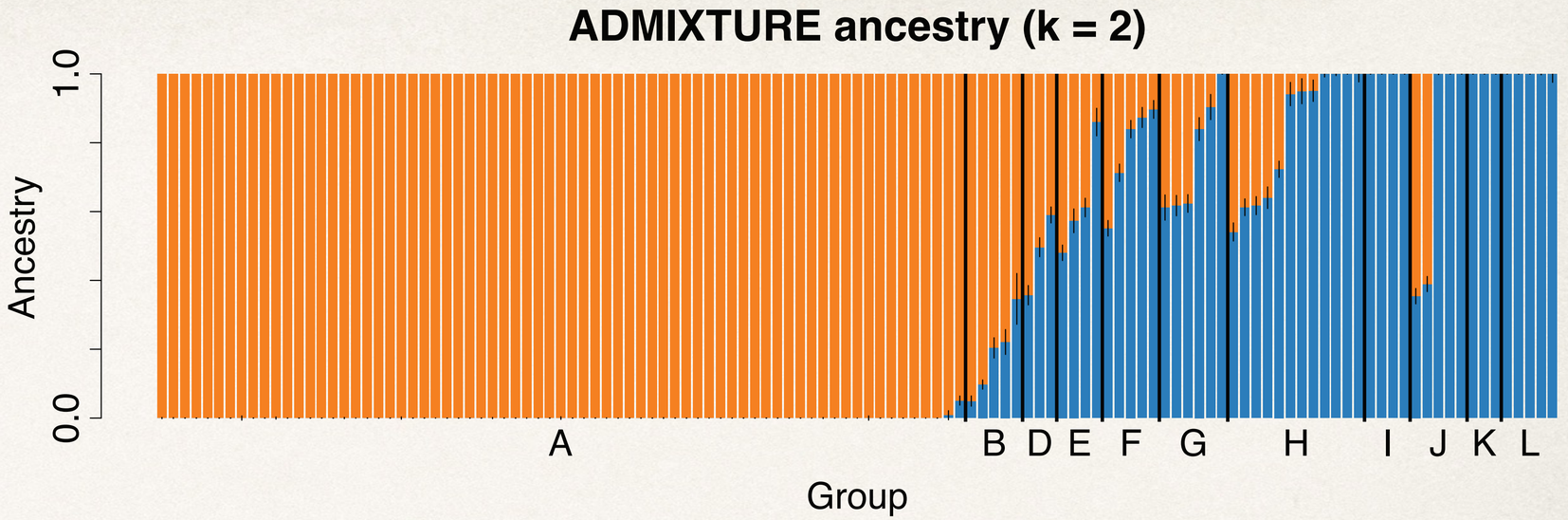
Cons

- *Not yet tested with whole genome sequencing*
- Not compatible with the mitochondrial genome
- Potential for interference from non-bacterial, CpG-methylated DNA
  - e.g., plant or animal prey, intestinal helminths, coprophagous animals?

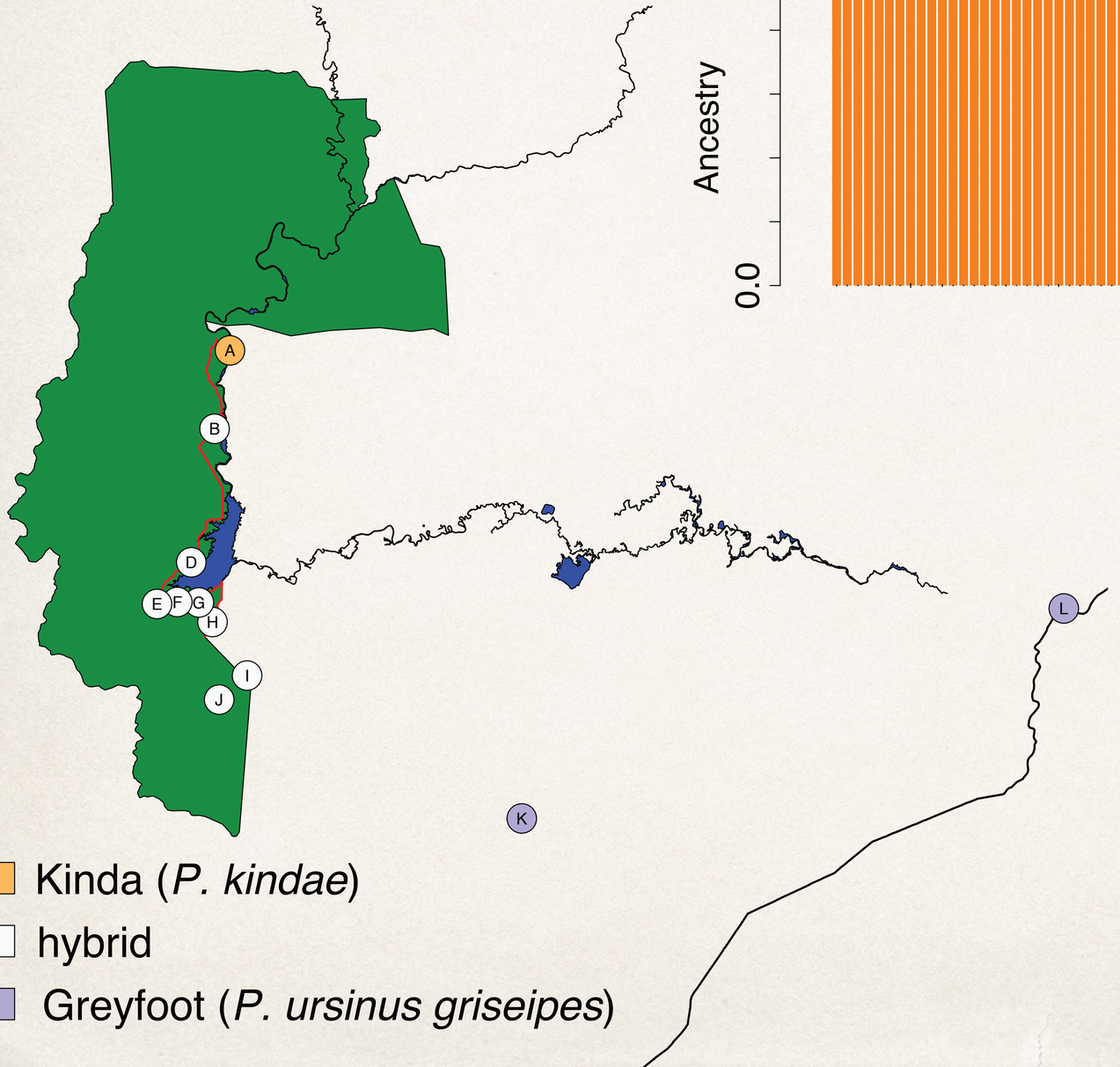
Concluding thoughts

- *Polishing a turd*
  - Primate DNA from feces is low-quantity
  - If the proportion is too low, it may not be worth the effort
  - If the absolute quantity is too low, expect dropout
- Our enrichment method is fast and effective with low bias
- With two options already, **noninvasive population genomics is a reality**

ongoing work with
J. Phillips-Conroy, C. Jolly, *et al.*



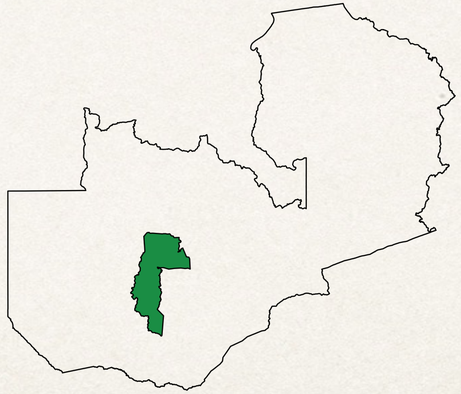
dataset with 14,642 SNPs



Kinda (*P. kindae*)

hybrid

Greyfoot (*P. ursinus griseipes*)



Kafue National Park,
Zambia

We'd like to hear from you!

Cold Spring Harbor Laboratory

bioRxiv beta

THE PREPRINT SERVER FOR BIOLOGY

New Results

FecalSeq: methylation-based enrichment for noninvasive population genomics from feces

Kenneth L Chiou, Christina M Bergey

doi: http://dx.doi.org/10.1101/032870

Acknowledgments

- Andy Burrell, Erbay Yigit, Jane Phillips-Conroy, Cliff Jolly, Todd Disotell
- New England Biolabs
- Zambia Wildlife Authority, University of Zambia School of Veterinary Medicine, Southwest National Primate Research Center
- National Science Foundation (BCS 1341018, BCS 1260816, BCS 1029302, SMA 1338524), NYU University Research Challenge Fund, Leakey Foundation, Wenner-Gren Foundation, National Geographic Society
- The NYU Genome Technology Center is funded by the National Institutes of Health (NIH/NCATS UL1 TR00038, NIH/NCI P30 CA016087)