Conservation Genetics Resources
https://doi.org/10.1007/s12686-021-01192-w

TECHNICAL NOTE

Genomic resources for an ecologically important rodent, Gunnison's prairie dogs (*Cynomys gunnisoni*)

Sydni Suzanne Joubran[1] · **Loren Cassin-Sackett**[1,2]

Received: 27 October 2020 / Accepted: 20 January 2021

Abstract
Genomic resources are under-developed for rodents, including well-studied species such as prairie dogs (*Cynomys* spp.). We conducted whole-genome resequencing on 10 Gunnison's prairie dogs (*C. gunnisoni*, GUPD) and identified 12,842,055 high-quality SNPs, from which four sets of bait sequences were created. We designed two sets each (containing either 20k or 60k baits) for projects using contemporary (120 bp baits) or historical (100 bp baits) DNA. These bait sets can be used to study a variety of ecological and evolutionary questions in prairie dogs and other ground squirrel taxa.

Keywords In-solution hybridization · Sequence capture · Marmotini · Ground squirrels

Genomic markers have enabled a mechanistic understanding of ecological and evolutionary processes (e.g., Jarvis et al. 2014; Fernández et al. 2018; Ahrens et al. 2019). However, these resources are distributed unevenly across the tree of life (Thomson and Shaffer 2010), being disproportionately available for charismatic megafauna (e.g., Cho et al. 2013; Zhao et al. 2013; Gordon et al. 2016). Still lagging are resources from the most diverse group of mammals—rodents—despite their ecological and evolutionary importance.

As keystone species of western North American grasslands, prairie dogs (*Cynomys* spp.) have been the subject of intensive research on disease ecology (e.g., Busch et al. 2013; Sackett et al. 2013; Eads 2014), social behavior (e.g., Hoogland 1982; Sugg et al. 1996), and communication (Kiriazis and Slobodchikoff 2006), and all five species are of conservation concern due to habitat loss, persecution and sylvatic plague (Hoogland 2001; Seglund et al., 2005). Prairie dogs exhibit extreme variation in environmental conditions (Castellanos-Morales et al. 2014, 2016) and ecological characteristics (Lehmer and Biggins 2005). Despite extensive study on prairie dogs, the first mitochondrial (Streich et al. 2019) and nuclear (Tsuchiya et al., 2020) genomes were sequenced just in the last year, and few other publicly available genetic resources exist (Jones et al. 2005; Sackett et al. 2009).

Using 10 previously collected tissue samples (Fig. 1a; Sackett et al. 2014) from across the Gunnison's prairie dog (*C. gunnisoni*; GUPD) species range, including 5 individuals from each subspecies, genomes were sequenced to 12× (5 samples) or 24× (5 samples; Fig. 1b) on an Illumina NovaSeq 6000 at Duke University Sequencing and Genomic Technologies. We retained all high-quality reads and aligned them to the *C. gunnisoni* reference genome (Tsuchiya et al. 2020) with BWA (Li and Durbin 2009) using the *mem* algorithm in paired-end mode. Bam files were sorted with Picardtools (http://broadinstitute.github.io/picard/) and indexed with samtools (Li et al. 2009).

Polymorphisms were identified and filtered in GATK v4 (DePristo et al. 2011) and vcftools v0.1.16 (Danecek et al. 2011) to include SNPs and multi-nucleotide polymorphisms (MNPs) with a base quality score >30 and a coverage depth >8 in each individual. This resulted in 12,842,055 polymorphisms (Table 1) across 12,628 scaffolds. A generalized linear model showed the number of SNPs per individual was best predicted by sequencing coverage ($p=0.001$) and non-significantly by geographic distance to the reference genome sampling location ($p=0.251$; Fig. 1b). We used vcftools to estimate the average number of heterozygous sites per sample, mean nucleotide diversity (pi), and the transition to transversion ratio (Table 1). Our quality filtering and SNP

✉ Loren Cassin-Sackett
cassin.sackett@gmail.com

[1] Department of Biology, University of Louisiana at Lafayette, 410 E. St. Mary Blvd., Lafayette, LA 70503, USA

[2] Department of Integrative Biology, University of South Florida, Tampa, FL 33620, USA

Published online: 05 February 2021

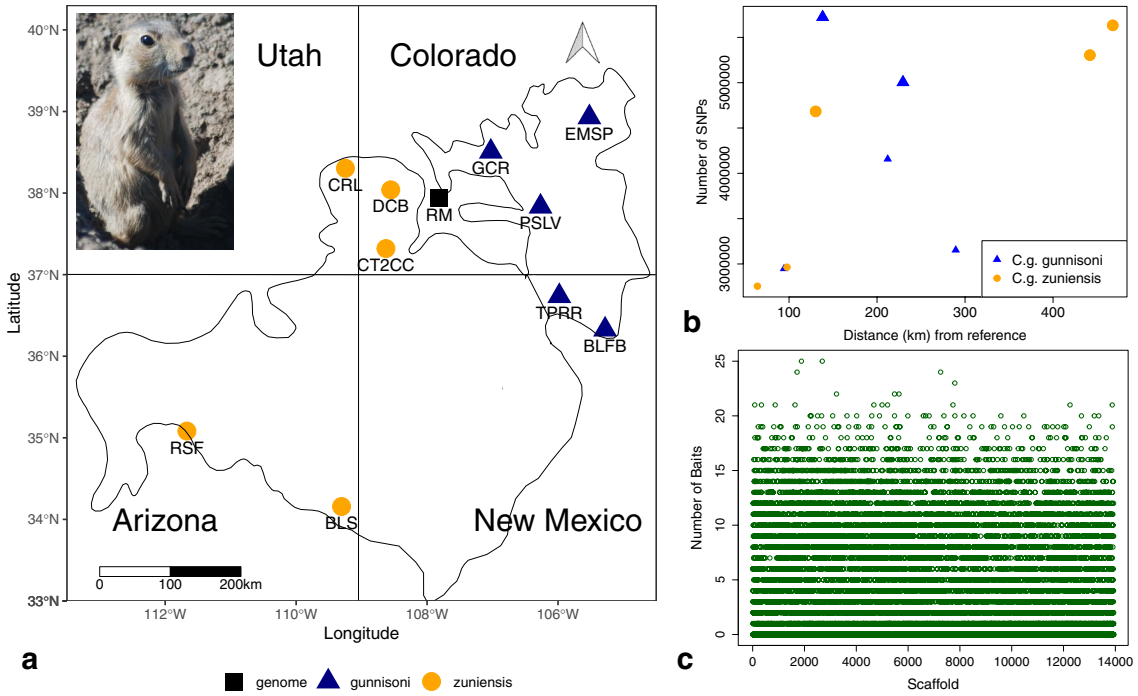

Fig. 1 **a** Map of locations in the Four Corners states (USA) for GUPD samples used in the development of the SNP array. Blue triangles represent *Cynomys gunnisoni gunnisoni*; orange circles represent *C. g. zuniensis*. The black square is the sampling location for the animal sequenced for the reference GUPD genome. More details about sampling locations can be found in Sackett et al. (2014). Inset: Gunnison's prairie dog in a low-elevation population in New Mexico. **b** Plot of the number of SNPs between each individual and the reference genome as a function of the geographic distance (in kilometers) between the two samples. Smaller points represent individuals resequenced to 12×coverage, while larger points are individuals resequenced to 24×coverage. **c** Number of baits per scaffold in the Bait Set 60k; each point represents the number of baits found on a single scaffold. The pattern is representative of coverage in other sets

Table 1 Summary statistics for SNPs identified among 10 GUPD and the reference genome

Summary statistics	
MNP	133,504
SNP	12,708,551
Transitions	8,404,676
Transversions	4,303,486
Transition:transversion ratio	1.953
Mean number heterozygous sites/sample	2,777,417.2
Pi	0.307552
Base substitutions	
A–C	1,112,735
A–G	4,203,331
A–T	1,169,792
C–G	909,098
C–T	4,201,345
G–T	1,111,861

MNPs multiple nucleotide polymorphisms, *SNPs* single nucleotide polymorphisms

selection parameters follow established standards (McKenna et al. 2010; Liu et al. 2012; Cassin-Sackett et al. 2019) and the pipeline, with specific parameters used in each step, is available on GitHub: https://github.com/CassinSackett/SNP_capture.

From these SNPs, we generated bait sets of different sizes, each corresponding to an upper limit for bait numbers (thereby maximizing cost efficiency) from Arbor Biosciences. We used BaitsTools (Campana 2018) to randomly select either 20,000 or 60,000 SNPs from across the genome. Bait sequences are complementary to specific regions of DNA and are used to target these regions for in-solution hybridization to DNA libraries. For use with contemporary DNA, we designed baits to be 120 base pairs in length (Cortes-Rodriguez et al. 2019) with the target SNP in the center of the sequence, requiring a minimum quality of 20 for all bases in the sequence. We scaled the number of allowable SNPs per scaffold to the scaffold length, enabling large scaffolds to produce more baits. Baits designed for use with historical DNA followed the same parameters except that bait length was limited to 100 bp as the fragmentation of historical DNA may result in decreased binding to longer baits. These shorter baits can also be used for modern samples (but might result in less specific binding). We removed all baits shorter than 80 base pairs in order to minimize

off-target reads due to the less-specific binding of shorter baits. Final bait sets included 19,999 ("Modern Baits 20k"), 60,000 ("Modern Baits 60k"), 19,997 ("Ancient Baits 20k"), and 59,996 baits ("Ancient Baits 60k"). As there was little overlap in SNPs between modern and ancient bait sets (2459 of 60k SNPs and 910 of 20k SNPs), for projects using both historical and contemporary DNA we recommend the 100 bp bait sets. Coverage of the genome by baits was relatively consistent across scaffolds (Fig. 1c). All four bait sets, as well as all 12,842,055 SNPs embedded in 120 bp baits, are available on FigShare (https://doi.org/10.6084/m9.figshare.11803326.v1).

Both bait sets are useful for genomics studies using neutral markers (Davidson et al. 2012; Giglio et al. 2020), whereas the 60k Bait Sets have denser sampling of the genome and are thus more appropriate for testing hypotheses about genomic regions under selection (Lowry et al. 2017). SNP arrays developed for one target species have been successfully applied to species ~ 37 MY divergent (Kharzinova et al. 2015; Minias et al. 2019); the radiation of *Marmotini* dates back only 7–9 MYA (McLean et al. 2018). Thus, our SNP arrays developed for GUPD have great promise for use not only in *Cynomys*, but also in > 90 other ground squirrel species.

Acknowledgements Computing was conducted on the high-performance cluster at the Louisiana Optical Network Initiative. Funding was provided by the University of South Florida and Colorado Parks and Wildlife. Samples were collected on historical Cheyenne, Hohokam, Hopi, Jicarilla Apache, Nʉmʉnʉʉ, Núu-agha-tʉvʉ-pʉ, Pueblos, Western Apache, Yavapai, and Zuni lands.

Funding Funding was provided by the University of South Florida and Colorado Parks and Wildlife. Bait sequences are available on FigShare (https://doi.org/10.6084/m9.figshare.11803326.v1); pipeline and scripts are available on GitHub: https://github.com/CassinSackett/SNP_capture.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

References

Ahrens CW, Mazanec RA, Paap T, Ruthrof KX, Challis A, Hardy G, Byrne M, Tissue DT, Rymer PD (2019) Adaptive variation for growth and resistance to a novel pathogen along climatic gradients in a foundation tree. Evol Appl 12(6):1178–1190

Busch JD, Van Andel R, Stone NE, Cobble KR, Nottingham R, Lee J, VerSteeg M, Corcoran J, Cordova J, Van Pelt W, Shuey MM, Foster JT, Schupp JM, Beckstrom-Sternberg S, Beckstrom-Sternberg J, Keim P, Smith S, Rodriguez-Ramos J, Williamson JL, Rocke TE, Wagner DM (2013) The innate immune response may be important for surviving plague in wild Gunnison's prairie dogs. J Wildl Dis 49(4):920–931

Campana MG (2018) BaitsTools: Software for hybridization capture bait design. Mol Ecol Resour 18(2):356–361

Cassin-Sackett L (2017) SNP_capture. GitHub. https://github.com/CassinSackett/SNP_capture.

Cassin-Sackett L, Callicrate TE, Fleischer RC (2019) Parallel evolution of gene classes, but not genes: evidence from Hawai'ian honeycreeper populations exposed to avian malaria. Mol Ecol 28:568–583

Castellanos-Morales G, Gasca-Pineda J, Ceballos G, Ortega J (2014) Genetic variation in a peripheral and declining population of black-tailed prairie dogs (*Cynomys ludovicianus*) from Mexico. J Mammal 95(3):467–479

Castellanos-Morales G, Gámez N, Castillo-Gámez RA, Eguiarte LE (2016) Peripatric speciation of an endemic species driven by Pleistocene climate change: the case of the Mexican prairie dog (*Cynomys mexicanus*). Mol Phylogenet Evol 94:171–181

Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, et al (2013) The tiger genome and comparative analysis with lion and snow leopard genomes. Nat Commun 4:2433

Cortes-Rodriguez N, Campana MG, Berry L, Faegre S, Derrickson SR, Ha RR, Dikow RB, Rutz C, Fleischer RC (2019) Population genomics and structure of the critically endangered Mariana Crow (*Corvus kubaryi*). Genes 10(3):187

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Lunter G, Marth G, Sherry ST, McVean G, Durban R (2011) The variant call format and VCFtools. Bioinformatics 27:2156

Davidson AD, Detling JK, Brown JH (2012) Ecological roles and conservation challenges of social, burrowing, herbivorous mammals in the world's grasslands. Front Ecol Environ 10(9):477–486

DePristo MA, Banks E, Poplin R, Garmella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498

Eads DA (2014) Factors affecting flea densities in prairie dog colonies: implications for the maintenance and spread of plague. Colorado State University

Fernández R, Kallal RJ, Dimitrov D, Ballesteros JA, Arnedo MA, Giribet G, Hormiga G (2018) Phylogenomics, diversification dynamics, and comparative transcriptomics across the spider tree of life. Curr Biol 28(9):1489-1497.e5

Giglio RM, Rocke TE, Osorio JE, Latch EK (2020) Characterizing patterns of genomics variation in the threatened Utah prairie dog: Implications for conservation and management. Evol Appl. https://doi.org/10.1111/eva.13179

Gordon D, Huddleston J, Chaisson MJP, Hill CM, Kronenberg ZN, Munson KM, et al (2016) Long-read sequence assembly of the gorilla genome. Science 352(6281):aae0344

Hoogland JL (1982) Prairie dogs avoid extreme inbreeding. Science 367(6483):1639–1641

Hoogland JL (2001) Black-tailed, gunnison's, and utah prairie dogs reproduce slowly. J Mammal 82(4):917–927

Jarvis ED, Ye C, Liang S, Yan Z, Zepeda ML, Campos PF, et al (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346(6215):1126–1138

Jones RT, Martin AP, Mitchell AJ, Collinge SK, Ray C (2005) Characterization of 14 polymorphic microsatellite markers for the black-tailed prairie dog (*Cynomys ludovicianus*). Mol Ecol Notes 5(1):71–73

Kharzinova VA, Sermyagin AA, Gladyr EA, Okhlopkov IM, Brem G, Zinovieva NA (2015) A study of applicability of SNP chips developed for bovine and ovine species to whole-genome analysis of reindeer rangifer tarandus. J Hered 106(6):758–761

Kiriazis J, Slobodchikoff CN (2006) Perceptual specificity in the alarm calls of Gunnison's prairie dogs. Behav Process 73:29

Lehmer EM, Biggins DE (2005) Variation in torpor patterns of free-ranging black-tailed and Utah prairie dogs across gradients of elevation. J Mammal 86(1):15–21

Li H, Durbin R (2009) Fast and accurate long-read alignment with Burrows-Wheeler Transform. Bioinformatics. 26:589

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, March G, Abecasis G, Durbin R (2009) The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079

Liu Q, Guo Y, Li J, Long J, Zhang B, Shyr Y (2012) Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data. BMC Genom 13:S8

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A (2017) Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour 17(2):142–152

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Research. 20:1297

McLean BS, Bell KC, Allen JM, Helgen KM, Cook JA (2018) Impacts of inference method and data set filtering on phylogenetic resolution in a rapid radiation of ground squirrels (Xerinae: Marmotini). Syst Biol 68(2):298–316

Minias P, Dunn PO, Whittingham LA, Johnson JA, Oyler-McCance SJ (2019) Evaluation of a Chicken 600K SNP genotyping array in non-model species of grouse. Sci Rep 9(1):6407

Picard Tools. Broad Institute, GitHub repository. http://broadinstitute.github.io/picard/

Sackett LC, Etchberger LK, Mazzella MN, Lim DD, Martin AP (2009) Characterization of 18 microsatellite loci for prairie dogs (genus *Cynomys*). Mol Ecol Resourc 10(1):232–236

Sackett LC, Collinge SK, Martin AP (2013) Do pathogens reduce genetic diversity of their hosts? Variable effects of sylvatic plague in black-tailed prairie dogs. Mol Ecol 22:2441–2455

Sackett LC, Seglund A, Guralnick RP, Mazzella MN, Wagner DM, Busch JD, Martin AP (2014) Evidence for two subspecies of Gunnison's prairie dogs (*Cynomys gunnisoni*), and the general importance of the subspecies concept. Biol Cons 174:1–11

Seglund AE, Ernst AE, Neill DMO (2005) Gunnison's prairie dog conservation assessment. Western Association of Fish and Wildlife Agencies

Streich SP, Keepers KG, Griffin KA, Kane NC, Martin AP (2019) The complete mitochondrial genome of Gunnison's prairie dog subspecies (*Cynomys gunnisoni gunnisoni*) and phylogenetic relationship within the genus *Cynomys*. Mitochondrial DNA Part B 4(1):397–398

Sugg DW, Chesser RK, Dobson FS, Hoogland JL (1996) Population genetics meets behavioral ecology. Trends Ecol Evol 8(11):338

Thomson RC, Shaffer HB (2010) Rapid progress on the vertebrate tree of life. BMC Biol 8:1–8

Tsuchiya M, Dikow R, Cassin-Sackett L (2020) First genome sequence of the Gunnison's prairie dog (*Cynomys gunnisoni*), a keystone species and player in the transmission of sylvatic plague. Mol Biol Evol 12:618

Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, et al (2013) Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nat Genet 45:67–71