Title: Empowering conservation practice with efficient and economical genotyping from poor quality samples using mPCRseq

Authors: MeghanaNatesh[1,2 #], Ryan W. Taylor[3,4,#], Nathan Truelove[5],Elizabeth A. Hadly[3], Stephen Palumbi[3,5], Uma Ramakrishnan[1#*] and Dmitri Petrov[3#*]

Author Affiliations:

[1]National Centre for Biological Sciences, TIFR, Bellary Road, Bangalore - 560065, India
[2]Sastra University, Tirumalaisamudram, Thanjavur – 613401, India

[3]Department of Biology, Stanford University, Stanford, CA - 94305, USA

[4]End2End Genomics LLC, Davis, CA

[5]Hopkins Marine Station, Stanford University, Pacific Grove, CA – 93950, USA

#: Equal contribution

email addresses: meghanan@ncbs.res.in, ryan@ryantaylor.net, truelovn@stanford.edu, hadly@stanford.edu, spalumbi@stanford.edu, uramakri@ncbs.res.in, dpetrov@stanford.edu

Running title: Genetic data to empower conservation

Keywords: conch, conservation genetics, endangered species, monitoring, new methods, next generation sequencing, non-invasive samples, SNPs, tigers

Type of article: Policy Perspectives

Number of words in the abstract:*196*
Number of words in the manuscript:*2548*
Number of references:*12*
Number of figures and tables: *Figure 3, Tables 0*

*Correspondence to: Uma Ramakrishnan, Associate Professor, Wellcome/DBT India Alliance Senior Fellow, National Centre for Biological Sciences, TIFR, Bangalore, 560065, India, +91-80-23666030, uramakri@ncbs.res.in
and Dmitri Petrov, Professor, Department of Biology, 371 Serra St., Stanford University Stanford, CA 94305-5020, USA, +1-(650) 736 1169, dpetrov@stanford.edu

Abstract: *Genetic tools are widely used in conservation but expansion to genomic scale data that are rapid and cost effective has been slow. Most genomic tools are developed for high quality DNA sources from lab or medical settings. So far, genetic data from market or field settings assess easily amplified mitochondrial DNA or a few microsatellites. Here we use multiplex PCR with low quality DNA from feces, hair, and cooked samples to quickly provide multi-locus data from many SNPs. We demonstrate the wide range of potential applications through tools to monitor individual wild tigers and track commercial trade in Caribbean queen conch. 100 SNPs from degraded tiger samples identified individuals, discerned close relatives, and detected population differentiation. 62 SNPs from conch fritters and field collected samples identified individuals, tested for close kin and detected population structure. Our study provides proof of concept for a rapid, simple, cost-effective, and scalable method, a framework that can be applied to other conservation scenarios previously limited by low quality DNA samples. These approaches provide a critical advance for wildlife monitoring and forensics, open the door to field-ready testing, and will strengthen the use of science in policy decisions and wildlife trade.*

50 **Introduction**

51 Stemming the tide of global species decline requires continuous monitoring and nimble,
52 adaptive management to promote species recovery. Effective monitoring requires, minimally,
53 the ability to identify and track presence of a species and ultimately to track specific individuals
54 and to determine their familial relationships. Integrating local data across species range
55 provides the ability to monitor global threats from population range reduction to illegal wildlife
56 trade.

57 In principle, all of these goals can be achieved via genotyping at a modest number of loci
58 such as microsatellites or single nucleotide polymorphisms (SNPs). Given that endangered
59 species tend to be rare and elusive, the necessary approach must be able to accommodate non-
60 invasive sources of DNA such as feces, shell, feathers, hair, and saliva that yield impure, mixed,
61 and/or extremely small amounts of degraded DNA. Moreover, market samples generated by
62 wildlife trade may be processed, cooked, dried, or mixed with other species, again providing
63 low quality and often mixed DNA. In contrast, the current approaches tend to require large
64 amounts of high quality, pure DNA from the target species most readily obtained by
65 immobilizing or sacrificing an animal (Fitak, Naidu, Thompson, & Culver, 2016; Kraus et al.,
66 2015) or demand expensive and generally inefficient enrichment strategies (Chiou & Bergey,
67 2018; Snyder-Mackler et al., 2016) to enhance the quantity of the target DNA. As a result, these
68 approaches are unsuitable, unaffordable, or too low throughput in most conservation settings.

69 The tiger (*Panthera tigris*) is one of the most endangered carnivores and despite drastic
70 habitat reduction, the remaining ~5,000 wild tigers inhabit over 14 countries. Their wide
71 distribution makes it critical that locally collected data is comparable across their range. While

72 photographic mark-recapture is commonly used to monitor tigers, capture rates are poor in
73 low-density habitats. Neither can photographs always be matched with confiscated tiger parts.
74 Instead, genotyping of scat or hair, along with rapid forensic testing of confiscated skins or
75 other traded parts can verify species, individual identity, and source populations.

76 Although formerly abundant, the queen conch (*Strombus gigas*) was listed in CITES
77 Appendix II in 1990. Trade was further restricted in 2004 due to declining stocks. Florida is
78 currently the largest market for queen conch seafood products, consuming 82% of international
79 trade. Matching the populations of origin of queen conch in Florida markets with international
80 regulations is crucial to recover this formerly lucrative fishery. The most direct access to
81 imported conch is from the hundreds of restaurants in Florida, necessitating techniques that
82 would allow genotyping from the most abundant menu item, fried conch fritters.

83 Here we demonstrate that a multiplex PCR approach (Campbell, Harmon, & Narum,
84 2015) followed by next-generation sequencing satisfies all the requirements necessary for
85 inexpensive, fast, and easy genotyping of low-quality samples. We illustrate the power of this
86 method for these two endangered species in very divergent conservation contexts and real-life
87 settings: genotypes from fecal, shed hair, and saliva samples found on the killed prey from wild
88 Indian tigers and from CITES-regulated Caribbean queen conch imported to the US and sold in
89 fried fritters.

90

91 **Methods**

92 We designed primers for 192 SNP loci from a pool of 50,000 SNPs discovered through
93 whole genome sequencing of 75 tigers of captive and wild origin from *P. t. tigris, P. t. jacksoni,*

94 *P. t. altaica, and P. t. sumatrae* subspecies (see supplementary materials). The pool consisted of
95 SNPs that were polymorphic in all subspecies (> 10% MAF), or differentiated subspecies from
96 each other (see supplementary materials). We used the publicly available program Primer3 to
97 design primer pairs (amplicon size 50-90 bp, primer size 17-25 bp, and Tm of 60-61ºC).The
98 resulting 192 primers were pooled for use in a single amplification of all target SNPs for each
99 tiger sample. The initial primer set was tested on fecal and tissue samples from captive tigers
100 (Table S1). The test set included fecal and tissue samples from the same individuals, multiple
101 subspecies, samples left to weather for up to two days, and a serially diluted tissue DNA
102 sample. Tiger DNA was quantified (see supplementary materials) and all samples were
103 genotyped in triplicate. SNPs were then tested on a range of non-invasive samples from across
104 India (representative photographs in Fig. 1b, sample list in Table S2). Multiple locations and
105 differing tiger landscapes, two sympatric carnivore species, the leopard (*Panthera pardus*) and
106 Indian wild dog or dhole *(Cuon alpinus*) were genotyped. Fecal samples, saliva and shed hair
107 were included.

108 The tiger example builds on a large number of SNPs ascertained from multiple fully
109 sequenced genomes. Often such resources are not available for endangered species, although
110 this is changing rapidly. For queen conch, we sought to identify SNPs from the transcriptomes
111 of 83 tissue samples from six populations in the Caribbean including Florida, collected from
112 fishing operations or non-lethally by research permit (see Table S3). Following methods
113 paralleling those used for tigers, we then genotyped at 192 SNPs an additional 237 conch (from
114 14 populations) and as well as 42 putatively conch samples isolated from 15 conch fritters
115 purchased at several Miami restaurants (see supplementary materials).

116 **Results**

117 SNP genotyping for wild tigers

118 Of the 192 SNPs, 126 produced the most consistent results (see supplementary
119 materials for details, Fig. S1, Table S4), and this panel of 126 SNP panel had a high overall
120 genotyping success rate (Fig. 2). An average of 95 SNPs (range: 4 - 114) were successfully typed
121 across all samples. When tiger DNA was > 0.01ng/ul (irrespective of sample type) an average of
122 105 SNPs (range: 48 - 114) were typed.12 SNPs did not amplify for over 50% of samples and
123 were excluded from subsequent analysis. The dhole sample had very poor amplification success
124 (mean across replicates <10 SNPs) as expected. Leopards had very high amplification success
125 (mean: 110 SNPs), but the SNPs were monomorphic and identical across two individuals (mean:
126 0.03% mismatches). All tiger samples in contrast had distinct genotypes and could be readily
127 distinguished from the leopard samples.

128 Replications of 'within-sample' and 'within individual-between tissue types' provided
129 important insights on the precision and accuracy of SNP identification and its efficacy for non-
130 invasive sampling. The proportion of matches between tissue types for the same individual
131 (n=5) was very high (precision, mean 0.97; range 0.91 - 0.99), showing that our approach
132 resulted in nearly identical genotypes for different sample types (feces, saliva, hair). Sample
133 replicates (within-sample comparisons) were also highly concordant and had a high proportion
134 of matches (precision, mean: 0.957, range: 0.757 - 1.0). Among these, 77 samples genotyped at
135 least 60% of SNPs had even higher concordance across replicates (mean proportion of matches:
136 0.969, range: 0.86-1.0), with only 5 samples showing values less than 0.95. These results are in
137 sharp contrast to the high error rates observed for microsatellite genotypes from other non-

138 invasive sampling techniques (Coates et al., 2011). Our probability of misidentifying two
139 different individuals as the same is vanishingly low (pID = 1.6E-22 assuming a population of
140 siblings).

141 Observed relatedness values (Fig. 3) and clustering of samples (Fig. S2) further highlight
142 our ability to detect biological patterns. Our data had two pairs of samples with known
143 relationships (zoo samples) while the remaining were sampled from different populations in the
144 wild. The known parent offspring and sibling pairs had relatedness values close to the expected
145 value of 0.5 (Figure 3). Observed pair-wise relatedness was higher within than between known
146 genetic clusters or populations (see Natesh et al., 2017). As expected, relatedness between
147 individuals from a small, isolated population (NW, Figure 3) was high. Even with the few
148 individuals in our pilot dataset presented here, wild individuals fell into three genetic
149 clusters/populations (Figure S2), similar to previous data (Natesh et al., 2017). For small
150 isolated populations (such as NW) future genotyping could help identify whether individuals are
151 also inbred, and if such inbreeding has fitness consequences through estimates of individual
152 reproductive success. Should genetic supplementation of this population be required, SNP data
153 from different tiger landscapes could help identify possible sources.

154

155 SNPs genotyping for queen conch

156 Our second example illustrates that this multiplex PCR approach can be applied even if
157 background genetic information is not as complete (as with tiger), and again highlights use in
158 difficult samples: in this case fried conch fritters. We identified 480,962 SNPs from the
159 sequenced transciptomes. Possibly because we based our primer design on a transcriptome

160 rather than a genome, our SNP success was lower for conch than tigers. Approximately half the
161 192 conch primer pairs failed to provide data. For the best 62 SNPs success was about the same
162 as for tigers (average 86%, Fig. S3). Replicate samples shared 99%-100% of their alleles, but we
163 also found several genetically identical samples among small processed conch pieces collected
164 from fishing boats. As with genetic re-sampling of tigers in the wild, discovery of repeated
165 individuals in a sample can inform overall fishery numbers. There were no obvious close
166 relatives among the samples, a result expected for low sample sizes and very large populations
167 (Fig. S3). The data also inform basic information important in design of marine protected areas.
168 Outlying islands (Aruba, St. Eustatius) were genetically differentiated from the central
169 Caribbean and Florida (F_{ST} = 0.037, 0.048 respectively, Table S5). Samples from the same island
170 group/on the same coast (e.g. Florida, Bahamas) were not differentiated. These patterns
171 parallel a recent survey of queen conch with microsatellites (Truelove et al., 2017).
172 While the conch isolated from deep fried fritters showed lower success rates than conch
173 sampled by fishermen or from biopsies, the success rates were high enough to allow individual
174 identification and initial population comparison. Twenty-three of the fritter conch did not
175 amplify with any primer, or produced data for less than 18 SNPs, and we suspect they were not
176 Queen conch. A further 8 fritter conch showed poor SNP amplification (average 41%) but could
177 be confirmed as Queen conch samples. The most successful 17 conch from fritters showed
178 success rates comparable to those of fresh samples (77% vs 86%) (Fig. S3). Even with this
179 preliminary data set, we can determine that the fritter samples showed lower (average) genetic
180 identity to Florida populations (average 78.6 – 80.5 alleles shared, Fig. S6)compared to the
181 Puerto Rico or St. Eustatius (81.1-82.3, Fig. S6). Positive identification of populations for

182 imported samples will require greater geographic sampling, a larger SNP panel, and a better
183 genotype map of potential source locations. For example, assignment testing successfully
184 places 65% of Bahamas conch samples back into that population, but this is not yet high enough
185 for precise population discovery. Nevertheless, we can already suggest that the Florida Keys
186 and Nassau are less likely to be the populations of origin for the fritters we genotyped, and
187 show that high resolution nuclear data can be readily obtained even from deeply processed
188 commercial samples and used to address key conservation needs.
189
190 **Discussion**
191 Our results on these small, pilot datasets provide proof of concept for multiplex SNP
192 genotyping of non-invasive and processed market samples from two species with vastly
193 different physiologies, biologies, ecologies, and conservation challenges. This approach satisfies
194 the key requirements for a practical conservation application. It is inexpensive to scale up, we
195 estimate that implementation costs (for upto 1000 SNPs) can be as low as $5 per sample. The
196 initial costs for primer design might include full-genome sequencing, which is becoming more
197 and more affordable. The cost of a genome of reasonable quality, even for fairly large
198 mammalian genomes, in the range of a few thousand dollars (Armstrong et al., 2017). In
199 addition, primers based on RADSeq data or RNAseq data provide cheaper solutions for very
200 large genomes. The largest development cost is of primer synthesis, which is expected to go
201 down as high-throughput oligo synthesis is becoming a large-scale, commercial commodity
202 service. Lab equipment required of the approach is minimal (centrifuge, thermocycler,
203 pipettes), and field-based sequencers are allowing the use of this method anywhere in the field.

204 Increasing the number of SNPs beyond a few hundred can provide additional
205 information. For instance, even with a few thousand SNPs it should be possible to start fully
206 imputing large regions of the genome in cases where full genomes of individuals in the pedigree
207 can be obtained. Moreover, strategic design of primers that would amplify clusters of closely
208 located SNPs should allow detection of very recent inbreeding by detecting long contiguous
209 runs of homozygosity (Kirin et al., 2010). Some of the SNPs could be located in such close
210 proximity that they generate minihaplotypes that have been shown to be particularly useful in
211 pedigree reconstruction (Baetscher, Clemento, Ng, Anderson, & Garza, 2018). The SNP panels
212 can be further designed to allow simultaneous species, individual, and diet identification for
213 clusters of species, such as, for instance, all large carnivores and all common species in their
214 diet in India or sub-Saharan Africa.
215 Most importantly, our data show that multiplex PCR is successful for degraded, cooked,
216 mixed, or small and low-concentration samples, making it an ideal tool for monitoring
217 individuals in populations under rigors of field conditions or commercial markets. In the future,
218 when populations maybe very small and isolated, we may need the ability to impute whole
219 genomes, construct full pedigrees, and conduct genome-wide association studies for
220 detrimental or advantageous traits. In principle, this method will allow such advances.
221 Effective monitoring of individuals, populations and species is critical to designing rapid
222 conservation action and management. In the future, conservation will depend critically on such
223 tools. Small, isolated populations will require inbreeding management, genotype tracking for
224 population assessment, and data to assess when individuals should be introduced. Populations
225 across fragmented landscapes may require assisted migration, especially in the context of

226 climate change and the range shifts it demands. Minimizing illegal wildlife trade will require
227 enforcement based on genetically identified sources of trafficked animals. Rapid genetic
228 monitoring of endangered species from commonly occurring non-invasive samples will provide
229 a data-pathway to species recovery. We believe that multiplex PCR presents an example of such
230 rapid, accessible, cheap and efficient technology that will make this possible.
231

232 **References**

233 Armstrong, E. E., Taylor, R. W., Prost, S., Blinston, P., Der, E. Van, Madzikanda, H., … Petrov, D.
234 (2017). Entering the era of conservation genomics: Cost-effective assembly of the African
235 wild dog genome using linked long reads. *BioRxiv*.
236 http://doi.org/http://dx.doi.org/10.1101/195180.
237 Baetscher, D. S., Clemento, A. J., Ng, T. C., Anderson, E. C., & Garza, J. C. (2018).
238 Microhaplotypes provide increased power from short-read DNA sequences for relationship
239 inference. *Molecular Ecology Resources*, *18*(2), 296–305. http://doi.org/10.1111/1755-
240 0998.12737
241 Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing
242 (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing.
243 *Molecular Ecology Resources*, *15*(4), 855–867. http://doi.org/10.1111/1755-0998.12357
244 Chiou, K. L., & Bergey, C. M. (2018). Methylation-based enrichment facilitates low-cost,
245 noninvasive genomic scale sequencing of populations from feces. *Scientific Reports*, *8*(1),
246 1–10. http://doi.org/10.1038/s41598-018-20427-9
247 Coates, B. S., Bayles, D. O., Wanner, K. W., Robertson, H. M., Hellmich, R. L., & Sappington, T.

248 W. (2011). The application and performance of single nucleotide polymorphism markers
249 for population genetic analyses of lepidoptera. *Frontiers in Genetics*, *2*(JUNE), 1–10.
250 http://doi.org/10.3389/fgene.2011.00038
251 Fitak, R. R., Naidu, A., Thompson, R. W., & Culver, M. (2016). A New Panel of SNP Markers for
252 the Individual Identification of North American Pumas. *Journal of Fish and Wildlife*
253 *Management*, *7*(1), 13–27. http://doi.org/10.3996/112014-JFWM-080
254 Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., Mckeigue, P. M., & Wilson, J. F. (2010).
255 Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE*,
256 *5*(11), 1–7. http://doi.org/10.1371/journal.pone.0013996
257 Kraus, R. H. S., vonHoldt, B., Cocchiararo, B., Harms, V., Bayerl, H., Kühn, R., … Nowak, C. (2015).
258 A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic
259 wolf monitoring in Europe based on noninvasively collected samples. *Molecular Ecology*
260 *Resources*, *15*(2), 295–305. http://doi.org/10.1111/1755-0998.12307
261 Natesh, M., Alta, G., Nigam, P., Jhala, Y. V, Zachariah, A., Borthakur, U., & Ramakrishnan, U.
262 (2017). Conservation priorities for endangered Indian tigers through a genomic lens.
263 *Scientific Reports*, *7*(1), 1–11. http://doi.org/10.1038/s41598-017-09748-3
264 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. a R., Bender, D., … Sham, P. C.
265 (2007). PLINK: a tool set for whole-genome association and population-based linkage
266 analyses. *American Journal of Human Genetics*, *81*(3), 559–75.
267 http://doi.org/10.1086/519795
268 Snyder-Mackler, N., Majoros, W. H., Yuan, M. L., Shaver, A. O., Gordon, J. B., Kopp, G. H., …
269 Tung, J. (2016). Efficient Genome-Wide Sequencing and Low Coverage Pedigree Analysis

from Non-invasively Collected Samples. *Genetics*, *203*(June), 699–714. http://doi.org/10.1534/genetics.116.187492

Truelove, N. K., Box, S. J., Aiken, K. A., Blythe-Mallett, A., Boman, E. M., Booker, C. J., … Stoner, A. W. (2017). Isolation by oceanic distance and spatial genetic structure in an overharvested international fishery. *Diversity and Distributions*, *23*(11), 1292–1300. http://doi.org/10.1111/ddi.12626

Acknowledgements: We thank San Diego Zoo Global, Conservation Society of California at Oakland Zoo, Oakland, CA, San Francisco Zoo, the Bronx Zoo, WCS Russia program, El Paso Zoo, Performing Animal Welfare Society, Wildlife Conservation Society India Program, Centre for Wildlife Studies, India, Wildlife Conservation Trust, Himanshu Chhattani, Anubhab Khan, Prachi Thatte, Aditya Joshi, Arun Zachariah, Erica Calcagno and Jackie Gai for tiger samples. We thank the Forest Departments of Rajasthan, Karnataka and Madhya Pradesh and the American Zoo Association, Tiger SSP for permissions. Awadhesh Pandit and NCBS next generation genomics facility helped with Illumina sequencing. We thank Steve Box, and Steve Canty for logistical help and support in framing conservation questions for conch. **Funding:** R43HG009482, Wildlife Conservation Trust grant to UR, Fulbright Nehru Academic exchange Program (UR) and Wellcome Trust/DBT India Alliance senior fellowship (UR), Summit Foundation and the Global Genome Initiative at the Smithsonian Institution (to NT and SRP and Steve Box) supported conch work. Author Contributions: MN, RWT, DP, UR, EAH, SP and NT designed the study. MN,

RWT and NT conducted lab work and analyses. MN, RWT, DP, UR, EAH, SP wrote the paper.

Competing interests: Ryan W. Taylor and End2End Genomics LLC have received NIH small business funding (R43HG009482) to develop molecular tools for the study of non-model species. The other authors declare that we have no competing interests. **Data Availability:** On publication, the data will be made available on the Dryad repository.

Figures

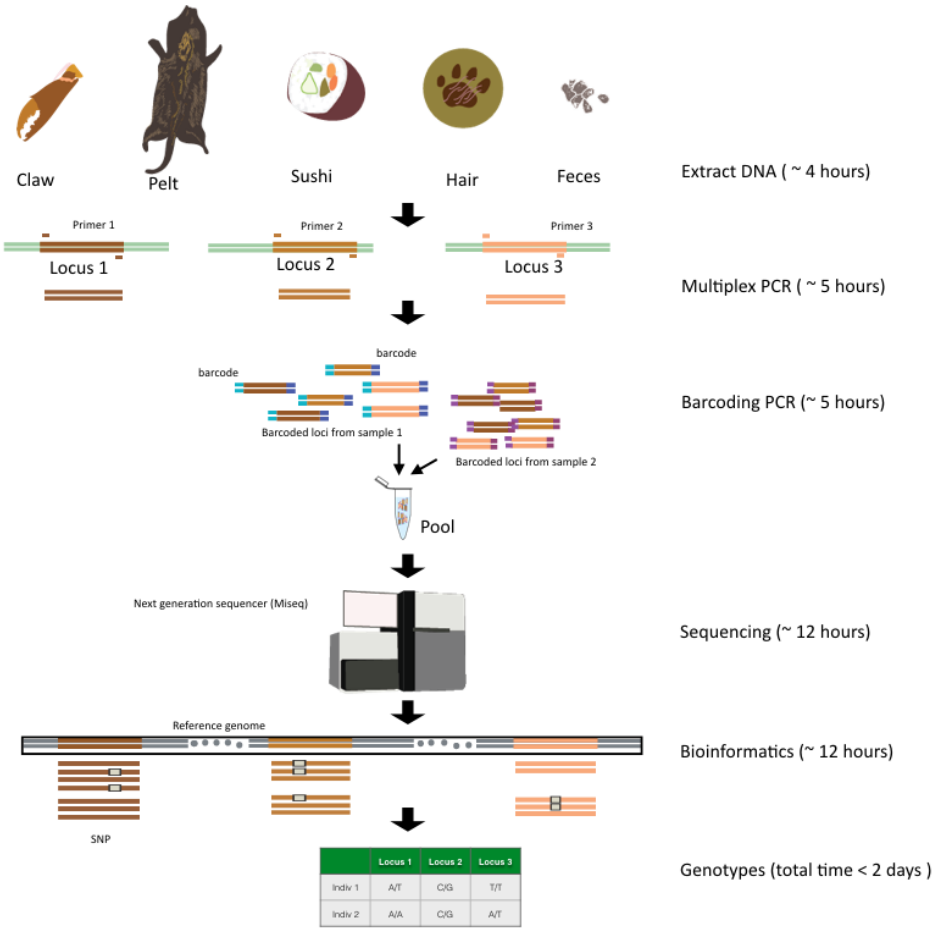

Figure 1a: A schematic showing the protocol and approximate time taken for each step.

Figure 1b: A panel of photographs showing the conservation setting for tiger and conch genomics. On the left, the top panel shows a tiger defecating (courtesy Himanshu Chattani), the middle panel a tiger scrape mark on a tree associated with hair samples (courtesy Kaushal Patel). On the right, the top panel shows a conch emerging from its shell and the middle a fisherman holding his catch.

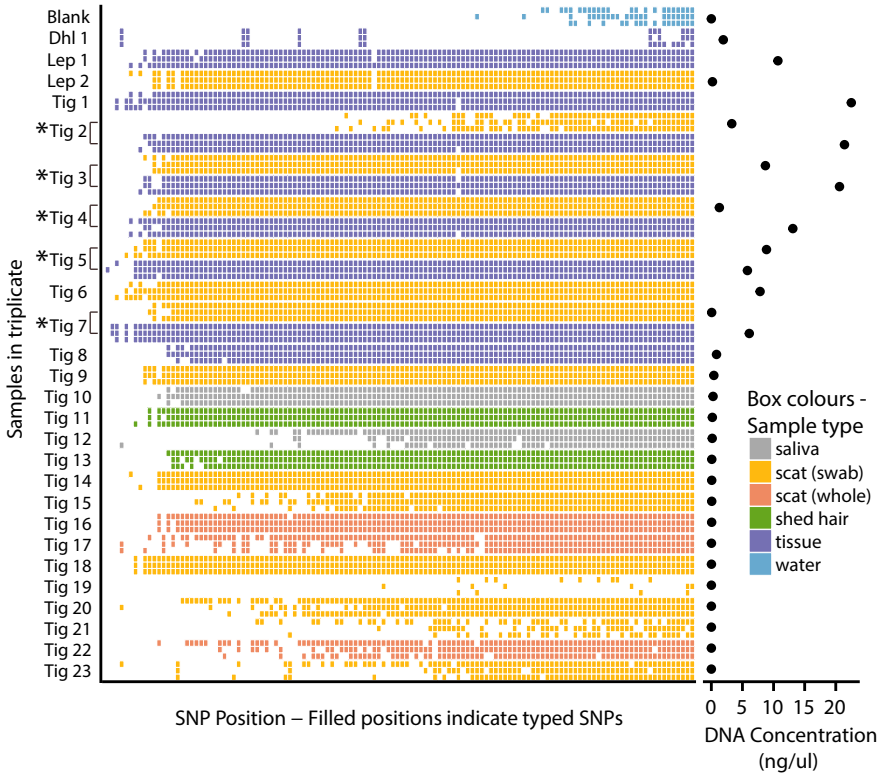

Figure 2: SNP typing success (in triplicate) for various tiger samples and controls. The concentration of each sample is shown on the right. Empty cells in a particular row correspond to SNPs that could not be genotyped. Filled cells indicate successfully typed SNPs, and the color indicates sample type.

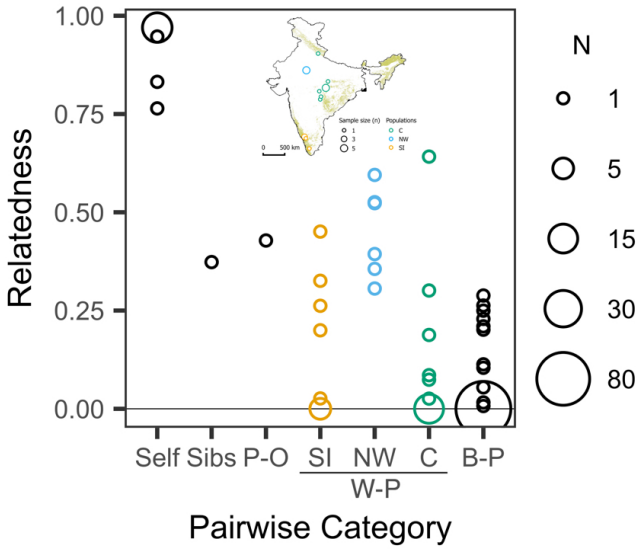

Figure 3: Pairwise relatedness (PI_HAT; (Purcell et al., 2007)) captures known relationships, and provides insights into wild tiger relatedness. 'Self' are pairwise comparisons among the three replicate analyses of each sample, as well as comparisons between different sample types for the same individual. Our dataset includes one pair of siblings (Sibs) and one parent-offspring pair (P-O). Sibs and P-O relationships were provided by zoo records. W-P are within population relatedness between samples from previously identified populations (Natesh et al., 2017). B-P are between population relatedness. Values in W-P and B-P are from field samples, or samples from wild caught captive tigers, where the relatedness is unknown. The inset shows sampling locations of the wild samples.