J. Indian Inst. Sci.
A Multidisciplinary Reviews Journal
ISSN: 0970-4140 Coden-JIISAD

REVIEW ARTICLE

Considerations for Initiating a Wildlife Genomics Research Project in South and South-East Asia

Anubhab Khan[1*] *and Abhinav Tyagi*[1,2]

Abstract | Next-generation sequencing (NGS) based genomic studies are revolutionizing the field of wildlife biology. These methods have yielded unprecedented insights for understanding ecology, evolution and conservation of wild populations. Despite the advantages, biodiversity rich regions in the tropics need more NGS-based studies to understand their native species. However, the field has progressed very slowly in these regions due to several challenges, including that most experts in the field are not based here. In this article, we highlight the factors that need to be considered before initiating a wildlife genomics research project with a focus on south and south-east Asia, though several factors apply to other regions as well. We highlight the challenges like policy issues for collecting samples and need for better sequencing and computational infrastructure. Finally, we discuss how global initiatives can help such regions setup NGS-based studies of wildlife.

1 Genomics in Conservation has Several Advantages over Traditional Genetics

The need for adopting genomics techniques to conservation research has been highlighted by several researchers[3, 16, 35]. The genetic data from NGS platforms generally yield more genetic markers than a typical microsatellite-based study. The obtained data is also less subjective than the microsatellite markers hence easy to share and more comparable across labs and easier to archive. The huge number of genetic markers that have been made accessible by the NGS platforms have changed the field of wildlife genetics. These include the ability to infer population wide phenomenon with few samples (for example demographic history[39, 58], recombination maps[13], inbreeding[19] and genetic variation[27]), increased confidence in genotypes called[73], ability to identify loci of functional importance[68, 88] and many others. Recently, genomic tools have availed other advantages like managing populations by identifying and ascertaining megafaunal species (for example tapanuli orangutan,[71], subspecies (tigers,[61], population structure (tigers,[72], understanding adaptations[8, 61], identifying genomic regions for planning genetic rescues that minimize adding deleterious alleles[87], rescue loci in runs of homozygosity[90] and hosting high frequency deleterious alleles [52]. SNP panels are being used to monitor wildlife[73,98] and introgression[75]. Apart from conservation, genomic tools have helped us to understand species evolution (for example canids,[38] and elephants,[78]), threats[63], management[40, 87], distribution and biogeography[108]. These insights would have been very difficult to achieve with traditional genetic markers. However, the majority of the species are studied in non-home range countries by non-native researchers.

2 Genomic Tools are New to Conservation but are Rapidly Being Adapted

The field of genomics started with the human genome project around the year 2000[37] and was subsequently adopted into wildlife studies, starting with the panda genome sequencing in the year 2010[59, 115]. Since then, genomic methods have been applied to study several wild species. The explosion in genomics research has been mainly triggered by development of new

[1] *National Centre for Biological Sciences, TIFR, Bangalore, India.*
[2] *SASTRA Deemed To Be University, Thanjavur, India.*
**anubhabkhan@gmail.com*

sequencing platforms, drop in sequencing costs and improvement in analysis pipelines[67]. The number of species with genomes assembled and sequenced have been rapidly increasing due to the contributions of laboratories mainly in North America and Europe. However, the technology has not been adopted widely in the developing countries. Most of this is probably because of the need for an advanced infrastructure for undertaking such research. Apart from the obvious requirement for next-generation sequencers, there needs to be temperature-controlled facilities for storing and archiving samples and super computers for data analysis, data storage, archival and retrieval. These are all generally expensive and less accessible especially for the poorly funded researchers in field of wildlife studies. South and south-east Asia is an example of such region where next-generation sequencing studies remains to be adopted widely in wildlife studies. This region is biodiversity rich and hosts many endemic and endangered species[96]. Facilities and easy access to latest technologies for genome sequencing are available in China, India and Singapore. In China, India and Singapore NGS tools have been used to study genetic variation (for example,[72], ascertain sub-species and population structure (for example,[8, 61], assemble genomes (for example,[65, 93] and more[14, 51, 73, 98]. Several countries in the region have limited or no access to high throughput sequencers engaged in research with more tangible economic outcomes (medicine and agriculture) but not for wildlife research. It is important to encourage genomics-based research in the biodiverse regions along the tropics. However, lack of exposure to the new techniques can impede the growth of the field. Here, we discuss the important considerations about planning a wildlife genomics research project. We primarily discuss DNA sequencing since this is more common than RNA sequencing and is a critical step in initiating genomic project on wildlife that do not have baseline genomic resources. We discuss the factors that need to be taken into account by researchers in this region regarding samples, library preparation, sequencing and analysis.

Box 1

Single nucleotide variant (SNV) and single nucleotide polymorphism (SNP)

SNVs are single nucleotide alleles caused by point mutations. To be called as a SNP, the SNV must be present in at-least 1% of the population[109].

Reference genome and genome assembly

Reference genome is a digital nucleic acid sequence, assembled as a representative set of genes in one idealized individual organism of a species. The term genome assembly refers to the process of assembling huge number of DNA sequences to create a representation of the original chromosomes (reference genome), from where the DNA originated[60].

Scaffolding

Scaffolding is the method of linking contiguous bases of DNA called contigs, assembled from many shorter reads, in the right order and orientation. The linking generally involves adding gaps represented by N[86].

Genome resequencing

Genome resequencing is method that utilizes sequencing the genome of individuals from a species for which a reference genome (same species or closely related one) is already available[99].

Optical and PCR duplicates

Duplicate reads in sequencing generally arise from sampling the exact same template DNA molecule. These are generally non-independent measurements of the DNA. These generally arise because of PCR (PCR duplicates) but also when a single cluster of reads is falsely used to compute two read cells separately. They are always present next to each other on the flow-cell (optical duplicates)[116].

Enrichment

It is a method of increasing the relative concentration of target nucleotide fragments to be sequenced[26].

3 Samples

The first prerequisite for wildlife genomics research is to obtain appropriate genetic material amenable to sequencing. Access to genetic material from wildlife is regulated in most countries and more so if the species are endangered[106]. Sampling of any kind of genetic material from endangered species requires special permits in most countries. These permits are granted by the managers/managing authority of the protected areas. It is important to know beforehand the kind of samples that will be required for the research questions to be addressed so that appropriate permissions can be obtained. In general, extracted RNA and DNA samples are used for sequencing. But certain types of questions require intact nuclei[21], histone bound DNA[79] and native

methylated DNA strands[94]. The sample types, their preservation, transport and nucleic acid extraction protocols depends on the questions being addressed and the availability of genomic resources. For example, intact nuclei and good quality RNA is needed for reference genome assembly and annotation while even degraded DNA can be used for genome resequencing.

3.1 Invasive Samples

Invasive sampling involves establishing physical contact with the individual and may cause discomfort to the animal[15, 23]. Such samples are very difficult to obtain permissions for from the authorities as it poses certain risks. For large bodied mammals, individuals may have to be tranquilized which can be potentially risky[2]. Such samples are generally obtained for small animals for example rodents, bats, lizards, small birds, insects and other species that may not have stringent protection[6, 18, 49]. At least within India, while rodents and lagomorphs can be sampled invasively under a proper ecological sampling framework[7, 29], large animals have to be sampled opportunistically[51]. Invasive sampling of plants may not face such challenges. Commonly invasive samples are blood samples obtained after capturing the individual[51]. Other methods include biopsy darts[15], ear and tail clippings[29] and fin clippings[114]. Such samples yield large amounts of high molecular weight DNA amenable to all kinds of library preparation and sequencing. If the research question demands RNA sequencing data of the species of interest, invasively collected tissue samples may be required[77].

Depending on the nucleic acid required, different kinds of preservation techniques are employed. One of the best methods is to collect the solid tissue in PBS solution (EDTA tubes for blood) then inoculate an appropriate cell culture medium with the invasively harvested tissue sample[20] and flash freeze the left-over tissue in liquid nitrogen. Such preservation methods can yield intact nuclei needed for certain applications[32]. However, many field sites are in remote locations with no access to electricity, no sterile rooms for cell culture and sometimes too remote to carry enough liquid nitrogen. In such cases preserving intact nuclei is a challenge that needs to be addressed. Presently, some of the best methods for preserving invasive samples at remote locations for NGS applications include collecting the tissue in absolute alcohol[31], Longmire's buffer[62], RNA later[22] and certain other proprietary solutions. For blood, EDTA tubes or PAXgene tubes work the best[25]. It is a good practice to store all of these samples in a freezer as soon as possible. If RNA is to be harvested from these samples, RNA later[22] and other RNA preserving solutions are needed that generally lyse the cells. RNA degrades faster than DNA due to its single stranded nature, the pentose sugar being more susceptible to hydrolysis and the stability of RNase enzymes. Hence they require delicate handling[36]. In some cases, isolating the nucleic acids as soon as possible and preserving the nucleic acids might work best.

Invasive samples have high endogenous nucleic acid content and hence, can be used directly for most sequencing applications. Presently, it is a logistical challenge for researchers in developing countries to maintain proper sample storage facilities at field sites due to lack of consistent electric supply, tropical climates leading to faster degradation, difficulty in maintaining cold chains during transport of samples from field site to lab, delay in importing reagents for immediate processing of samples, lack of field-based laboratories and the general costs involved.

3.2 Non-invasive Samples

Individuals can also be sampled non-invasively. It is relatively easier to obtain permissions to collect such samples as individuals are not disturbed. These samples can be collected in a proper ecological sampling framework and have been utilized for census studies[98], estimating pedigrees[51], predicting disease risk[54]. Majority of wildlife studies involving large bodied animals utilize such samples. These include feces[48, 105, 104], shed hair[51], excretory products[42], shed antlers[44], shed feathers[45], dandruff[30], scales[56], pellets[29], saliva[112] and corpses[51]. These samples may not yield RNA but DNA can be reliably extracted. However, for many endangered species non-invasive samples may be the only option.

The major challenges in sequencing DNA from non-invasive samples is the fragmentation of DNA[84], low quantities of endogenous DNA[51], difficulty in lysis of samples[51] and abundance of exogenous DNA[51]. However, methods that rely on amplicon sequencing[73], enrichment of DNA (Box 1)[26], enrichment of cells[76] and discarding parts more likely to be contaminated[64] can tackle many of these challenges. For large carnivores at least non-invasive samples are encountered more often than invasive samples[51]. This may allow for shorter sampling periods than invasive samples[51]. It might also be easier to preserve and transport certain non-invasive samples. While most

non-invasive samples may not be suited for establishing cell lines, samples like feathers, hair, antlers and scales can be transported and preserved without specialized buffers. For other samples, preserving in lysis buffers, RNA later or alcohol[53] might be better. For long term preservation, it is better to transport and store these samples at low temperatures.

Non-invasive samples are not optimal for applications like genome assembly or annotation. Additionally most of these samples contain high proportions of nucleases from non-endogenous cells[83], contain high numbers of bacteria and fungi that can degrade nucleic acids and the samples could have been exposed to harsh environmental conditions before collection[118]. This leads to lower success rates in certain sequencing applications. The challenges faced in transport and processing of certain non-invasive samples are the same as those for invasive samples. Samples like shed hair, shed antler, shed feathers and bones from corpses maybe transported in room temperature[70].

3.3 Environmental DNA

Environmental DNA (eDNA) is also gaining in popularity[17]. eDNA can be obtained from soil, water, parasites and others. However, these samples have been used mostly to obtain DNA for species identification[17]. These approaches have been useful in estimating the biodiversity, species surveys, phylogenetics, and environmental impact assessments. The most common environmental DNA samples are soil and water[12, 41]. However, sampling of parasites like leeches[92] and tsetse flies[34] are increasing in popularity. These are easier to obtain permissions for in most countries. All eDNA may be considered as non-invasive samples, however, unlike other non-invasive samples it is often challenging to perform population genetic analysis with these samples[1].

The major challenge in processing eDNA is enrichment[102]. Here, enrichment might be needed to first concentrate the material that is usable for sequencing since the volume of sample collected is high but DNA content is low. Most methods involve precipitating cells and nucleic acids or manually removing unwanted parts. Generally mitochondrial DNA is easier to obtain from these due to a higher proportion of mitochondria in cells than DNA in the nucleus. The DNA obtained is expected to be fragmented and in low concentration. The challenges faced in the preservation and transport of these samples is similar to that of non-invasive samples.

4 Library Preparation

The DNA that is obtained from the various samples has to be prepared for sequencing since the raw DNA is not suitable for sequencing directly and several modifications are needed. These are of course dependent on the questions one aims to address in their study and the type of sequencer to be used. Here, we discuss the most common reasons for sequencing genomes of wild species on a next-generation sequencing platform and the most common types of library preparations.

4.1 Reference Genome Assembly

Reference genomes are extremely useful in all genomics analysis. They are generally needed for ascertaining the genomic position of a sequencing read. This in turn is essential for identifying SNPs, identifying consequences of mutations, identifying arrangement of genes and many others[28, 82]. There are two kinds of assemblies, the most common being draft genome assembly where the genome is sequenced and only a preliminary assembly is done generally without scaffolding and annotating[28]. The other is complete genome assembly where the genome is sequenced, assembled by scaffolding and then annotated[28]. Complete genome assemblies might be rare for non-model wild species however initiatives like earth biogenome project[57], zoonomia consortium[117] and DNA zoo (https://www.dnazoo.org) are making several high-quality genomes available. Such initiatives generally help obtain an assembled genome of closely related species which can be useful if the genome of the species of interest is not available.

Assembling a reference genome generally involves sequencing across multiple sequencing platforms for long and short read data. The primary assembly software often relies on the assumption that the sequencing reads are from a randomly fragmented genome[55, 101]. Hence, it is important that the starting DNA is not fragmented or contaminated and is of high concentration. PCR free library preparations are preferred for assembly hence high amounts of DNA is needed[55]. All these conditions are generally met by DNA from invasively collected tissue or solid tissue from fresh corpses.

Several platforms are available for reference genome assembly. Until recently one of the easiest and cheapest option was to obtain a 10 × genomics chromium library for a good quality primary assembly[8] but this is presently unavailable. Other option for good primary assembly includes short read data from DNA libraries of varying insert

sizes but these can be expensive. These primary assemblies can be scaffolded using Bionano optical reads, Dovetail Chicago libraries and Dovetail HiC libraries[33]. These have yielded chromosome level assemblies. However, there will be several gaps in the assembly which can be filled with long read sequences. All of this is generally very expensive ($30,000–$40,000 USD for mammalian genomes) and needs advanced computing facilities. Such resources are generally rare for wildlife research in south and south-east Asian countries. Here, global genome assembly initiatives[57, 117] can be very helpful. However, species that do not survive in captivity and have to be sampled from wild populations are difficult to obtain invasive samples from. Here, it is important for researchers from native range countries of these species to participate in the global initiatives to develop resources for studying the species of interest further.

4.2 Whole Genome Resequencing

Whole genome sequence data can provide unprecedented insights into the biology of a species. Whole genome sequences have been used to resolve subspecies[8, 61], gain insights into species extinctions[69], understand wildlife disease[68], demographic history and ongoing threats[63], develop population management strategies ([52]), discover new species[71], develop tools to study the current species and many other applications. Analysis of whole genome sequencing data requires the availability of a reference genome assembly of the species or a closely related species.

Most common whole genome re-sequencing projects make use of short read data. This is due to the availability of PCR-based library preparations which make the fragmented and low concentration DNA obtained from non-invasive samples amenable to whole genome sequencing. Other techniques involve enriching the host DNA from non-invasive samples before whole genome sequencing[26, 76, 80]. Presently, enzyme-based DNA fragmentation and tagging (for example those implemented in kits like Illumina DNA prep) and PCR-based library preparations (for example those implemented in NEB Ultra 2 library preparation kits) are very useful in sequencing genomes from very small amounts of DNA obtained from non-invasive samples[8, 51]. Analysis of whole genome sequences requires that the sequencing reads belong to randomly fragmented DNA. This is generally difficult to know for non-invasive samples and PCR-based methods might bias this further. Hence, it is crucial to remove reads that are optical duplicates (Box 1) from the data before further analysis.

Whole genome resequencing projects might require very high-throughput sequencers which are rare in many south and south-east Asian countries. The sequencing can be expensive too ($1000–$1500 for mammalian genomes including library preparation and sequencing at 20 × depth). However, library preparations, when done in-house rather than outsourced to dedicated sequencing companies, can be cheap and only require common equipment. This can allow researchers to ship their libraries to neighbouring countries for sequencing. Local sequencing hubs in the region are China, India and Singapore with access to all library preparation and sequencing technologies.

4.3 Reduced Representation Sequencing

For several studies data from few hundred to few thousand genomic loci tend to be sufficient. These involve coarse grained studies on population structure, individual identification, phylogenetics, pedigree and many others. The genomic regions sequenced depend on the study design. These methods generally make analysis of large sized genomes easier by ignoring repetitive sites of the genomes[43]. Reduced representation of genome sequence (RRGS) approaches includes exome capture[74], genotyping by sequencing[97], ultra conserved elements[95] and transcriptome sequencing[11].

Reduced representation sequencing is a robust and cost-effective approach to generate genome wide data[103] and is based on the second generation of high-throughput DNA sequencing technology[9]. Single nucleotide polymorphisms (SNPs) are the predominant genomic data sought by wildlife biologists. Hence, approaches like Restriction site Associated DNA Sequencing (RADSeq) and double digest RADSeq (ddRAD-Seq) are common[10, 81]. Ultra-conserved elements are sequenced predominantly for phylogenetics research, while transcriptome and exome are sequenced primarily for understanding adaptations[85] and functional changes[111] for population genetics[113] and ecology[50, 46].

The major advantage of reduced representation sequencing is that it can generate the desired data from poor quality biological samples like non-invasive samples such as fecal and hair samples[26], Tyagi et al. in prep). A reference genome is desired for the same or related species to analyse the obtained data for identification of SNP markers. It is more crucial to have

Table 1: Common library preparations and applications in example fields of study.

Nucleic acid	Samples used	Library preparation	Enrichment required	Fields of study
DNA	Non-invasive sample	Whole genome	Yes	Population genetics, phylogenetics, evolution
DNA	Invasive	Whole genome	No	Genome assembly, population genetics, phylogenetics, evolution
DNA	Non-invasive sample	RAD	Yes	Population genetics
DNA	Invasive	RAD	No	Population genetics
DNA	Non-invasive sample	Baiting	Yes	Population genetics
DNA	Invasive, non-invasive sample	Amplicon	No	Population genetics, population census
DNA	Non-invasive sample	UCE	Yes	Phylogenetics
DNA	Invasive	Bisulphite	No	Functional genetics
DNA	Invasive, non-invasive sample	Exome	Yes	Population genetics, phylogenetics, evolution
RNA	Invasive	Transcriptome	No	Population genetics, phylogenetics, evolution, functional genetics

a reference genome if the sample type is poor like non-invasively collected samples. For non-invasive samples, the reference genome not only helps in arranging and sorting the sequencing reads but is also useful for filtering out the non-endogenous and non-target DNA such as those from environment or diet. However, one of the biggest advantages is the availability of analytical programmes and pipelines which enable the discovery of SNPs without a reference genome like stacks[24]. Most reagents and facilities needed for this type of library preparation are available in most south and south-east Asian countries and basic computational resources may be sufficient for analysing this type of data. Table 1 lists the most common types of library preparations and the types of samples they are suitable for.

5 Sequencing

It is important to know beforehand the purpose of sequencing and gathering as much information as available about the genome of the species. The few important issues to know beforehand are whether a reference genome for the species is available and if not then if the reference genome of a closely related species is available. It is also useful to know the genome size and the GC content of a genome, as these are strong predictors of how difficult sequencing, assembly or analysis may be. These information are very important in deciding the type of sequencing to be used and for deciding a budget. Depending on the questions and the resources available the type of sequencing data to be obtained is chosen. Majorly there are short read sequencing and long read sequencing. Short read sequencing is the most common, easy to analyse and cheapest kind of data available presently while long read sequencing is generally expensive, has fewer analysis pipeline and the sequencing facilities may be rare. Arumugam et al.[9] have listed several sequencing platforms and their advantages.

5.1 Short Read Sequencing

This type of sequencing generally yields sequencing reads of lengths between 100 and 250 base pairs. Sequencing can be done either as single reads where each DNA fragment in the library is sequenced from one side or as paired reads where each DNA fragment is sequenced from both ends and information on pairing of reads is retained. Presently this is the most common type of sequencing data used for most species as it is cheap (less than $1000 for sequencing mammalian genomes at 20×) and easy to analyze. Presently, Illumina sequencing platforms are most commonly used for this sequencing. Previously, Roche 454 was an option, but has been discontinued. Within south-east Asia, presently China, India, Indonesia, Philippines, Singapore and Thailand are potentially the countries where large scale illumina sequencing platforms (Illumina *HiSeqX10* and above) are available, Bangladesh, Malaysia and Vietnam have medium

scale sequencers (Illumina *HiSeq 4000* etc.) while Nepal and Myanmar have small scale sequencers (Illumina *MiSeq*) available. Researchers in several countries like Bhutan, Brunei, Cambodia and Laos may not have access to any next-generation sequencing platforms. Researchers in these countries have to obtain CITES permits to take their samples to other countries for sequencing.

Short read sequencing data is presently important for almost all sequencing applications including reference genome assembly and annotation, resequencing, transcriptome sequencing, reduced representation sequencing and others. For non-invasive and environmental samples this kind of sequencing might be the only viable option since the DNA is generally fragmented. This type of sequencing is also suitable for low concentration DNA due to the availability of PCR-based library preparation techniques.

5.2 Long Read Sequencing

This type of sequencing yields long sequencing reads of lengths 10 Kbp and above or ultra-long sequencing reads going up to 1 Mb or more. This type of sequencing can be quite expensive (more than $5000 for sequencing mammalian genomes at 20 ×) and difficult to analyze. Generally, longer read lengths have higher error rates and high sequencing depths are needed to correct for these. Some ultra-long read sequencers have non-random errors which means sequencing on a different platform is needed to correct for these. Long read sequencing is commonly done on PacBio sequencing platforms while Oxford Nanopore sequencers are increasing in popularity. Presently, Oxford Nanopore is probably the only sequencing platform that is reported to provide ultralong sequencing reads ranging upto few Mbp in read length[4, 47]. Pacbio sequencers are available in Bangladesh, China, India, Malaysia, Philippines, Singapore, Thailand and Vietnam. Oxford nanopore also have a portable device that is commonly available in several countries or can be obtained easily however, the *promethion*, a high-throughput device is rare in most of south and south-east Asia. Pacbio sequencers claim to have the most unbiased random sequencing errors while Oxford nanopore might have non-random sequencing errors[91].

Long read sequencing is most commonly used for applications involving genome assembly. They have not been used for resequencing projects for wildlife studies. This kind of sequencing needs good quality and high amounts of DNA like those obtained from invasive samples. Long read sequencing library preparations are mostly free of several rounds of PCR.

Box 2

Video URL for understanding common sequencing technologies

Illumina: https://youtu.be/fCd6B5HRaZ8

PacBio: https://youtu.be/_lD8JyAbwEo

Nanopore: https://youtu.be/RcP85JHLmnI, https://youtu.be/sv9fFeSd3kE

Ion Torrent: https://youtu.be/zBPKj0mMcDg

Video URL for understanding few scaffolding technologies

HiC: https://youtu.be/tTjmRtgnNng, https://youtu.be/Ba8O6rSoU8A

Chicago: https://youtu.be/tTjmRtgnNng

Bionano: https://youtu.be/S2ng6glu04I

6 Analysis

This is a critical part of any scientific question to be addressed. For next generation sequencing-based experiments, obtaining data is relatively easy and a major portion of the work consists of analysing the data. Since NGS experiments can generate huge amounts of data in a short span of time, good computational infrastructure is needed to analyze it. The analysis of the data can be divided into filtering the data, stratifying the data, making inferences.

Filtering of the data is needed at various stages of the analysis to increase the accuracy of the results. The first step in any sequencing analysis involves the filtering of the raw data. This discarding involves removing adapters and indexes leftover after sequencing, trimming the bases towards the end of sequencing reads to increase the average quality of the reads and any reads that may be too short, are discarded as their genomic origins are difficult to determine accurately[100] and references therein). After this initial round of filtering the data may need further filtering down the pipeline to remove reads that maybe PCR duplicates in whole genome-based studies or to remove reads whose position along the genome cannot be ascertained confidently.

The filtered raw data needs to be clustered for further analysis. The reads are randomly output by the sequencing machine. It is important to determine the position of the bases either with respect to other bases in the sequencing data or with respect to a reference genome. This can be ascertained either by assembling the data into

genomes, aligning the data to a reference genome, aligning the reads with other reads to obtain relative positions or using targeted loci where the genomic location of the reads is known beforehand. Once the data has been stratified certain bins may be discarded due to low complexity or low depth depending on the questions being asked. This further allows for comparison of data from one region to another or data from one individual to another thus allowing for quantification of several biological parameters.

Lastly inferences need to be made from the data to finally address the questions. This often involves identifying SNPs, comparing gene sequences, computing phylogenies, quantifying population genetic parameters and other analysis depending on the questions.

Most of the softwares used for analysing and drawing inferences from NGS data are freely available and there is lots of support from peers for most of these. Platforms like github (https://github.com) are popular for sharing codes and scripts while biostars (https://www.biostars.org) and stackoverflow (https://stackoverflow.com) are being used commonly for discussing technical issues. However, computational costs remain one of the major constraints. Huge amounts of processing power and memory is needed for analysing NGS data. For most mammalian whole genome analysis, a minimum of 10 core processor with 128 Gb random access memory would be needed costing about $2500. This might be sufficient for data from most reduced representation library sequences. However, this minimum configuration may not enable multiple lab members to run their analysis at the same time and better computers would be needed. Additionally, for multi-individual range wide analysis or for genome assembly-based analysis much higher computing power is needed. Also, since the volume of data is very high, lots of storage space is needed. This can further add to the costs. For most NGS-based studies access to high throughput computing clusters and supercomputers might be essential. Several countries in south and south-east Asia lack such infrastructure for wildlife studies while this might be available for other use. China, India and Singapore might be the only hubs in the region with all the infrastructure for wildlife genomics studies. It is also easy to share computational resources across countries through remote access. Additionally, servers like galaxy[89] are designed specifically for such tasks.

7 Discussion

Here we discuss the major considerations to keep in mind for conducting wildlife genomics research in south and south-east Asia. Wildlife genomics studies based out of this region have been generally rare and most have been in collaboration with western labs. Initiating independent wildlife genomics studies in this region needs to take into account policy, funding, infrastructure and expertise. Presently in this region, for questions regarding biodiversity assessments amplicon sequencing from environmental DNA would be the easiest to begin since this is easier to obtain permissions for sampling, mostly involves amplifying and sequencing specific regions of genome and can be analysed with minimal computational infrastructure (for example[5, 107]. The reference panels for specific genomic regions are generally available on NCBI (https://www.ncbi.nlm.nih.gov) or easy to create. For questions regarding population genetic analysis within species obtaining non-invasive samples might be the easiest. Depending on the question and the sample, either a reduced representation library or a whole genome library can be prepared and sequenced. This however might need an available reference genome from the same or related species for analysis. The computational infrastructure needed can be good workstations or high throughput clusters depending on the data. The data can be archived on free online databases like SRA (https://www.ncbi.nlm.nih.gov/sra) (for example[51, 98]. The most difficult initiative for most countries in this region might be to address questions that require genome assemblies. These require invasive samples, several kinds of library preparations and whole genome sequencing on various platforms and high throughput computational clusters (for example[93]. Several developments and collaborations are needed to initiate such studies in most countries. Presently the disparity in technology in various country is a major problem that needs to be tackled for a holistic growth across the globe.

There needs to be policy changes in the region to advance wildlife biology. There need to be policies for collecting and archiving of tissue samples from corpses of species, creation of cell lines, archiving of non-invasively collected samples and accessing these collections. The corpses of elusive species are extremely rare and yield extremely good quality samples for most studies[51]. Presently these samples are collected for forensic analysis only and rarely archived for research. Additionally, large museum facilities are absent in this region. The region needs to urgently develop

museums like Field Museum of Chicago or Natural History Museum, London where all the existing biodiversity can be archived properly with care. Such facilities not just help tackle the sampling problems in wildlife genomic studies but also are consistent with scientific principles where the results are repeatable. There needs to be a major policy overhaul to implement these.

Funding seems to be the next major challenge in the region. While several countries in the region are rich and invest in scientific research, most of the countries are not very invested in wildlife genetics research that does not yield immediate tangible benefits. Several non-governmental organizations have been providing small grants to support field work but very few to none provide the financial resources needed to undertake independent wildlife genomics research. The labs in these regions need support from the government, foreign organizations like US Fish and Wildlife Services and from independent donors to establish wildlife genomics research locally. An even challenging task is to support long term monitoring projects which are almost non-existent in the region for the majority of endangered species. Initiatives like Isle of Rum deer project, meerkats, Serengeti lions are desperately needed for several endemic and endangered species in this species rich region. This will need committed funding or specialized funding agencies from the government. International consortiums and initiatives like earth biogenome project[57] and Zoonomia consortium[117] will hopefully help in improving the situation. Hopefully, if policies can be implemented the funding situation might improve too. Websites like https://terravivagrants.org list a few funding sources for wildlife research that might aid in acquiring funds for this region.

While certain infrastructure needed for undertaking wildlife genomics projects exist in most of the countries already, they may not be accessible to wildlife researchers. For example, most supercomputers or clusters might be dedicated for defence and monitoring purposes while most sequencers might be dedicated for medical or agricultural research. Facilities for large scale storage and archiving of samples for centralized access are rare. Transporting samples from a remote field site to the laboratory is a challenge since maintaining a cold chain is very difficult and field-based laboratory facilities are rare. If infrastructural challenges in sampling can be overcome, the library preparation facilities reach this part of the world very late probably due to low demand. Most of the library preparation and sequencing technology is developed in the western countries and there can be a lag of a few years before these technologies are available in south and south-east Asia. This further delay the genomics research in this region.

Presently, most genomics research in the region appears to be dependent on collaborations with western laboratories. These generally involves transferring samples from native country of a species to labs abroad. This is highly regulated by international treaties and acts like the CITES[110], the Nagoya protocol[66] and other local regulations. Also, while researching in foreign labs allows local researchers to get trained, most researchers may not run independent labs in the region to address wildlife genomics questions. This might be because of lack of funding and infrastructure. Building local capacities by global collaborations will not only aid greatly in developing the field also for tackling global crises like pandemics and global warming locally and in parallel. Once these are made available hopefully the countries will start attracting their local talents and lots of wildlife genomics research can happen in the native range of the species.

Publisher's Note

Acknowledgements

We thank Chui Li for the information on availability of various sequencers across the region.

Declarations

Conflict of Interest

The authors declare that they have no conflict of interest.

Received: 18 February 2021 Accepted: 17 May 2021
Published online: 3 July 2021

References

1. Adams CI, Knapp M, Gemmell NJ, Jeunen GJ, Bunce M, Lamare MD, Taylor HR (2019) Beyond biodiversity: can environmental DNA (eDNA) cut it as a population genetics tool? Genes 10(3):192
2. Ahmed J, Buragohain N, Mekola I, Kyarong S, Choudhury B, Ahmed N (2020) First extant record of Royal Bengal Tiger (Panthera tigris) in Dibang valley of Arunachal Pradesh, India with a note on translocation

using Xylazine and ketamine anaesthetics. J Entomol Zool Stud 8(2):531–533
3. Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11(10):697–709
4. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q (2020) Opportunities and challenges in long-read sequencing data analysis. Genome Biol 21(1):1–16
5. Andriyono S, Alam MJ, Kim HW (2019) Environmental DNA (eDNA) metabarcoding: diversity study around the Pondok Dadap fish landing station, Malang, Indonesia. Biodiversitas 20(12):3772–3781
6. Angelier F, Tonra CM, Holberton RL, Marra PP (2010) How to capture wild passerine species to study baseline corticosterone levels. J Ornithol 151(2):415–422
7. Ansil BR, Mendenhall IH, Ramakrishnan U (2021) High prevalence and diversity of Bartonella in small mammals from the biodiverse Western Ghats. PLoS Negl Trop Dis 15(3):e0009178
8. Armstrong E, Khan A, Taylor RW, Gouy A, Greenbaum G, Thiéry A, Kang JT, Redondo SA, Prost S, Barsh G, Kaelin C (2019) Recent evolutionary history of tigers highlights contrasting roles of genetic drift and selection. Mol Biol Evol 38(6):2366–2379
9. Arumugam R, Uli JE, Annavi G (2019) A review of the application of next generation sequencing (NGS) in wild terrestrial vertebrate research. Annu Res Rev Biol 31(5):1–9
10. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376. https://doi.org/10.1371/journal.pone.0003376
11. Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J 51(5):910–918
12. Barnes MA, Turner CR, Jerde CL, Renshaw MA, Chadderton WL, Lodge DM (2014) Environmental conditions influence eDNA persistence in aquatic systems. Environ Sci Technol 48(3):1819–1827
13. Barroso GV, Puzović N, Dutheil JY (2019) Inference of recombination maps from a single pair of genomes and its application to ancient samples. PLoS Genet 15(11):e1008449
14. Baveja P, Garg KM, Chattopadhyay B, Sadanandan KR, Prawiradilaga DM, Yuda P, Lee JG, Rheindt FE (2021) Using historical genome-wide DNA to unravel the confused taxonomy in a songbird lineage that is extinct in the wild. Evol Appl 14(3):698–709
15. Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. Mol Ecol Resour 9(5):1279–1301
16. Benestan LM, Ferchaud AL, Hohenlohe PA, Garner BA, Naylor GJ, Baums IB, Schwartz MK, Kelley JL, Luikart G (2016) Conservation genomics of natural and managed populations: building a conceptual and practical framework. Mol Ecol 25(13):2967–2977
17. Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Douglas WY, De Bruyn M (2014) Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol Evol 29(6):358–367
18. Brown C (2007) Blood sample collection in lizards. Lab Anim 36(8):23–25
19. Brüniche-Olsen A, Kellner KF, Anderson CJ, DeWoody JA (2018) Runs of homozygosity have utility in mammalian conservation and evolutionary studies. Conserv Genet 19(6):1295–1307
20. Burkard M, Whitworth D, Schirmer K, Nash SB (2015) Establishment of the first humpback whale fibroblast cell lines and their application in chemical risk assessment. Aquat Toxicol 167:240–247
21. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J (2013) Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol 31(12):1119–1125
22. Camacho-Sanchez M, Burraco P, Gomez-Mestre I, Leonard JA (2013) Preservation of RNA and DNA from mammal samples under field conditions. Mol Ecol Resour 13(4):663–673
23. Carroll EL, Bruford MW, DeWoody JA, Leroy G, Strand A, Waits L, Wang J (2018) Genetic and genomic monitoring with minimally invasive sampling methods. Evol Appl 11(7):1094–1119
24. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124-f3140
25. Chai V, Vassilakos A, Lee Y, Wright JA, Young AH (2005) Optimization of the PAXgene^{TM} blood RNA extraction system for gene expression analysis of clinical samples. J Clin Lab Anal 19(5):182–188
26. Chiou KL, Bergey CM (2018) Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. Sci Rep 8(1):1–10
27. Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, Oh S, Kim HM, Jho S, Kim S, Shin YA (2013) The tiger genome and comparative analysis with lion and snow leopard genomes. Nat Commun 4:2433
28. Church DM, Schneider VA, Graves T, Auger K, Cunningham F, Bouk N, Chen HC, Agarwala R, McLaren WM, Ritchie GR, Albracht D (2011) Modernizing reference genome assemblies. PLoS Biol 9(7):e1001091
29. Dahal N, Kumar S, Noon BR, Nayak R, Lama RP, Ramakrishnan U (2020) The role of geography, environment, and genetic divergence on the distribution of pikas in the Himalaya. Ecol Evol 10(3):1539–1551
30. Dennis C (2006) Conservation at a distance: a gentle way to age. Nature 442(7102):507–509
31. Derkarabetian S, Benavides LR, Giribet G (2019) Sequence capture phylogenomics of historical

ethanol-preserved museum specimens: Unlocking the rest of the vault. Mol Ecol Resour 19(6):1531–1544

32. Elbers JP, Rogers MF, Perelman PL, Proskuryakova AA, Serdyukova NA, Johnson WE, Horin P, Corander J, Murphy D, Burger PA (2019) Improving illumina assemblies with Hi-C and long reads: an example with the North African dromedary. Mol Ecol Resour 19(4):1015–1026
33. Field MA, Rosen BD, Dudchenko O, Chan EK, Minoche AE, Edwards RJ, Barton K, Lyons RJ, Tuipulotu DE, Hayes VM, Omer D, A. (2020) Canfam_GSD: De novo chromosome-length genome assembly of the German Shepherd Dog (Canis lupus familiaris) using a combination of long reads, optical mapping, and Hi-C. GigaScience 9(4):giaa027
34. Gaithuma A, Yamagishi J, Hayashida K, Kawai N, Namangala B, Sugimoto C (2020) Blood meal sources and bacterial microbiome diversity in wild-caught tsetse flies. Sci Rep 10(1):1–10
35. Garner BA, Hand BK, Amish SJ, Bernatchez L, Foster JT, Miller KM, Morin PA, Narum SR, O'Brien SJ, Roffler G, Templin WD (2016) Genomics in conservation: case studies and bridging the gap between data and application. Trends Ecol Evol 31(2):81–83
36. Gayral P, Weinert L, Chiari Y, Tsagkogeorga G, Ballenghien M, Galtier N (2011) Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. Mol Ecol Resour 11(4):650–661
37. Giani AM, Gallo GR, Gianfranceschi L, Formenti G (2020) Long walk to genomics: History and current approaches to genome sequencing and assembly. Comput Struct Biotechnol J 18:9–19
38. Gopalakrishnan S, Sinding MHS, Ramos-Madrigal J, Niemann J, Castruita JAS, Vieira FG, Carøe C, de Manuel Montero M, Kuderna L, Serres A, González-Basallote VM (2018) Interspecific gene flow shaped the evolution of the genus Canis. Curr Biol 28(21):3441–3449
39. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. Nat Genet 43(10):1031
40. Grossen C, Guillaume F, Keller LF, Croll D (2020) Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nat Commun 11(1):1–12
41. Harrison JB, Sunday JM, Rogers SM (2019) Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proc R Soc B 286(1915):20191409
42. Hedmark E, Flagstad Ø, Segerström P, Persson J, Landa A, Ellegren H (2004) DNA-based individual and sex identification from wolverine (Gulo gulo) faeces and urine. Conserv Genet 5(3):405–410
43. Hirsch CD, Evans J, Buell CR, Hirsch CN (2014) Reduced representation approaches to interrogate genome diversity in large repetitive plant genomes. Brief Funct Genomics 13:257–267
44. Hoffmann GS, Johannesen J, Griebeler EM (2015) Species cross-amplification, identification and genetic variation of 17 species of deer (Cervidae) with microsatellite and mitochondrial DNA from antlers. Mol Biol Rep 42:1059–1067
45. Hogan FE, Cooke R, Burridge CP, Norman JA (2008) Optimizing the use of shed feathers for genetic analysis. Mol Ecol Resour 8:561–567
46. Hölzer M (2021) A decade of de novo transcriptome assembly: are we there yet? Mol Ecol Resour 21(1):11–13
47. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, Malla S (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 36(4):338–345
48. Joshi A, Vaidyanathan S, Mondol S, Edgaonkar A, Ramakrishnan U (2013) Connectivity of Tiger (*Panthera tigris*) populations in the human-influenced forest mosaic of Central India. PLoS ONE 8(11):e77980
49. Kawamichi T, Liu J (1990) Capturing and marking pikas (Ochotona) with systematic ear clipping patterns. J Mammal Soc Jpn 15(1):39–43
50. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B (2014) The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol 12(6):e1001889
51. Khan A, Patel K, Bhattacharjee S, Sharma S, Chugani AN, Sivaraman K, Hosawad V, Sahu YK, Reddy GV, Ramakrishnan U (2020) Are shed hair genomes the most effective noninvasive resource for estimating relationships in the wild? Ecol Evol 10(11):4583–4594
52. Khan A, Patel K, Shukla H, Viswanathan A, van der Valk T, Borthakur U, Nigam P, Zachariah A, Jhala Y, Kardos M, Ramakrishnan U (2021) Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indiantigers. bioRxiv
53. Kilpatrick CW (2002) Noncryogenic preservation of mammalian tissues for DNA extraction: an assessment of storage methods. Biochem Genet 40(1):53–62
54. Kim GS, Kim TS, Son JS, Lai VD, Park JE, Wang SJ, Jheong WH, Mo IP (2019) The difference of detection rate of avian influenza virus in the wild bird surveillance using various methods. J Vet Sci 20(5):e56
55. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ (2009) Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+ C)-biased genomes. Nat Methods 6(4):291–295
56. Kumar VP, Rajpoot A, Shukla M, Kumar D, Goyal SP (2016) Illegal trade of Indian Pangolin (Manis

crassicaudata): genetic study from scales based on mitochondrial genes. Egypt J Forensic Sci 6:524–533
57. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, Durbin R, Edwards SV, Forest F, Gilbert MTP, Goldstein MM (2018) Earth BioGenome Project: sequencing life for the future of life. Proc Natl Acad Sci 115(17):4325–4333
58. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. Nature 475(7357):493–496
59. Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, Zhang Z (2010) The sequence and de novo assembly of the giant panda genome. Nature 463(7279):311–317
60. Li J, de Vries RP, Peng M (2020) Reference module in life science, encyclopedia of mycology
61. Liu YC, Sun X, Driscoll C, Miquelle DG, Xu X, Martelli P, Uphyrkina O, Smith JL, O'Brien SJ, Luo SJ (2018) Genome-wide evolutionary analysis of natural history and adaptation in the world's tigers. Curr Biol 28(23):3840–3849
62. Longmire JL, Maltbie M, Baker RJ (1997) Use of "lysis buffer" in DNA isolation and its implication for museum collections. Museum of Texas Tech University, 163, pp 1–4
63. de Manuel M, Barnett R, Sandoval-Velasco M, Yamaguchi N, Vieira FG, Mendoza MLZ, Liu S, Martin MD, Sinding MHS, Mak SS, Carøe C (2020) The evolutionary history of extinct and living lions. Proc Natl Acad Sci 117(20):10927–10934
64. Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, Tomsho LP, Packard MD, Zhao F, Sher A, Tikhonov A (2008) Sequencing the nuclear genome of the extinct woolly mammoth. Nature 456(7220):387–390
65. Mittal P, Jaiswal SK, Vijay N, Saxena R, Sharma VK (2019) Comparative analysis of corrected tiger genome provides clues to its neuronal evolution. Sci Rep 9(1):1–11
66. Morgera E, Tsioumani E, Buck M (2014) Unraveling the nagoya protocol: a commentary on the nagoya protocol on access and benefit-sharing to the convention on biological diversity. Martinus Nijhoff Publishers
67. Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, Zhang J, Weinstock GM, Isaacs F, Rozowsky J, Gerstein M (2016) The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biol 17(1):1–9
68. Murchison EP, Schulz-Trieglaff OB, Ning Z, Alexandrov LB, Bauer MJ, Fu B, Hims M, Ding Z, Ivakhno S, Stewart C, Ng BL (2012) Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. Cell 148(4):780–791
69. Murray GG, Soares AE, Novak BJ, Schaefer NK, Cahill JA, Baker AJ, Demboski JR, Doll A, Da Fonseca RR, Fulton TL, Gilbert MTP (2017) Natural selection shaped the rise and fall of passenger pigeon genomic diversity. Science 358(6365):951–954
70. Nagy ZT (2010) A hands-on overview of tissue preservation methods for molecular genetic analyses. Org Divers Evol 10:91–105
71. Nater A, Mattle-Greminger MP, Nurcahyo A, Nowak MG, De Manuel M, Desai T, Groves C, Pybus M, Sonay TB, Roos C, Lameira AR (2017) Morphometric, behavioral, and genomic evidence for a new orangutan species. Curr Biol 27(22):3487–3498
72. Natesh M, Atla G, Nigam P, Jhala YV, Zachariah A, Borthakur U, Ramakrishnan U (2017) Conservation priorities for endangered Indian tigers through a genomic lens. Sci Rep 7(1):1–11
73. Natesh M, Taylor RW, Truelove NK, Hadly EA, Palumbi SR, Petrov DA, Ramakrishnan U (2019) Empowering conservation practice with efficient and economical genotyping from poor quality samples. Methods Ecol Evol 10(6):853–859
74. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461:272–276
75. Oliveira R, Randi E, Mattucci F, Kurushima JD, Lyons LA, Alves PC (2015) Toward a genome-wide approach for detecting hybrids: informative SNPs to detect introgression between domestic cats and European wildcats (Felis silvestris). Heredity 115(3):195–205
76. Orkin JD, Montague MJ, Tejada-Martinez D, de Manuel M, Del Campo J, Hernandez SC, Di Fiore A, Fontsere C, Hodgson JA, Janiak MC, Kuderna LF (2021) The genomics of ecological flexibility, large brains, and long lives in capuchin monkeys revealed with fecalFACS. Proc Natl Acad Sci 118(7):e2010632118
77. Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12(2):87–98
78. Palkopoulou E, Lipson M, Mallick S, Nielsen S, Rohland N, Baleka S, Karpinski E, Ivancevic AM, To TH, Kortschak RD, Raison JM (2018) A comprehensive genomic history of extinct and living elephants. Proc Natl Acad Sci 115(11):E2566–E2574
79. Park PJ (2009) ChIP–seq: advantages and challenges of a maturing technology. Nat Rev Genet 10(10):669–680
80. Perry GH, Marioni JC, Melsted P, Gilad Y (2010) Genomic-scale capture and sequencing of endogenous DNA from feces. Mol Ecol 19(24):5332–5344
81. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135
82. Pop M, Phillippy A, Delcher AL, Salzberg SL (2004) Comparative genome assembly. Brief Bioinform 5(3):237–248

83. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto J-M, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD, Wang J (2010) A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59–65
84. Reddy PA, Bhavanishankar M, Bhagavatula J, Harika K, Mahla RS, Shivaji S (2012) Improved methods of carnivore faecal sample preservation, DNA extraction and quantification for accurate genotyping of wild tigers. PLoS ONE 7(10):e46732
85. Rellstab C, Dauphin B, Zoller S, Brodbeck S, Gugerli F (2019) Using transcriptome sequencing and pooled exome capture to study local adaptation in the giga-genome of Pinus cembra. Mol Ecol Resour 19(2):536–551
86. Rice ES, Green RE (2019) Annu Rev Anim Biosci 7:17–40
87. Robinson JA, Räikkönen J, Vucetich LM, Vucetich JA, Peterson RO, Lohmueller KE, Wayne RK (2019) Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. Sci Adv 5(5):eaau0757
88. Ryder O, Miller W, Ralls K, Ballou JD, Steiner CC, Mitelberg A, Romanov MN, Chemnick LG, Mace M, Schuster S (2016) Whole genome sequencing of California condors is now utilized for guiding genetic management. In: International plant and animal genome XXIV conference, San Diego, CA, 8–13 January 2016
89. Saif R, Ejaz A, Mehmood T, Asif F, Alghanem SM, Ahmad TS (2020) Introduction to galaxy platform for NGS variant calling pipeline. Adv Life Sci 7(3):129–134
90. Saremi NF, Supple MA, Byrne A, Cahill JA, Coutinho LL, Dalén L, Figueiró HV, Johnson WE, Milne HJ, O'Brien SJ, O'Connell B (2019) Puma genomes from North and South America provide insights into the genomic consequences of inbreeding. Nat Commun 10(1):1–10
91. Scheunert A, Dorfner M, Lingl T, Oberprieler C (2020) Can we use it? On the utility of de novo and reference-based assembly of Nanopore data for plant plastome sequencing. PLoS ONE 15(3):e0226234
92. Schnell IB, Thomsen PF, Wilkinson N, Rasmussen M, Jensen LR, Willerslev E, Bertelsen MF, Gilbert MTP (2012) Screening mammal biodiversity using DNA from leeches. Curr Biol 22(8):R262–R263
93. Shingate P, Ravi V, Prasad A, Tay BH, Garg KM, Chattopadhyay B, Yap LM, Rheindt FE, Venkatesh B (2020) Chromosome-level assembly of the horseshoe crab genome provides insights into its genome evolution. Nat Commun 11(1):1–13
94. Smith ZD, Gu H, Bock C, Gnirke A, Meissner A (2009) High-throughput bisulfite sequencing in mammalian genomes. Methods 48(3):226–232
95. Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT (2014) Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst Biol 63(1):83–95
96. Sodhi NS, Koh LP, Brook BW, Ng PK (2004) Southeast Asian biodiversity: an impending disaster. Trends Ecol Evol 19(12):654–660
97. Spindel J, Wright M, Chen C, Cobb J, Gage J, Harrington S, Lorieux M, Ahmadi N, McCouch S (2013) Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor Appl Genet 126:2699–2716
98. Srivathsa A, Rodrigues RG, Toh KB, Zachariah A, Taylor RW, Oli MK, Ramakrishnan U (2021) The truth about scats and dogs: Next-generation sequencing and spatial capture–recapture models offer opportunities for conservation monitoring of an endangered social canid. Biol Conserv 256:109028
99. Stratton M (2008) Nat Biotechnol 26:65–66
100. Sun K (2020) Ktrim: an extra-fast and accurate adapter-and quality-trimmer for sequencing data. Bioinformatics 36(11):3561–3562
101. Sundquist A, Ronaghi M, Tang H, Pevzner P, Batzoglou S (2007) Whole-genome sequencing and assembly with high-throughput, short-read technologies. PLoS ONE 2(5):e484
102. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol 21(8):2045–2050
103. Van Tassell CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods 5(3):247–252
104. Thatte P, Chandramouli A, Tyagi A, Patel K, Baro P, Chhattani H, Ramakrishnan U (2020) Human footprint differentially impacts genetic connectivity of four wide-ranging mammals in a fragmented landscape. Divers Distrib 26:299–314
105. Thatte P, Joshi A, Vaidyanathan S, Landguth E, Ramakrishnan U (2018) Maintaining tiger connectivity and minimizing extinction into the next century: insights from landscape genetics and spatially-explicit simulations. Biol Conserv 218:181–191

106. Tsioumani T, Morgera E (2010) Wildlife legislation and the empowerment of the poor in Asia and Oceania. FAO legal papers online 83:1–124
107. Utzeri VJ, Schiavo G, Ribani A, Tinarelli S, Bertolini F, Bovo S, Fontanesi L (2018) Entomological signatures in honey: an environmental DNA metabarcoding approach can disclose information on plant-sucking insects in agricultural and forest landscapes. Sci Rep 8(1):1–13
108. Vianna JA, Fernandes FA, Frugone MJ, Figueiró HV, Pertierra LR, Noll D, Bi K, Wang-Claypool CY, Lowther A, Parker P, Le Bohec C (2020) Genome-wide analyses reveal drivers of penguin diversification. Proc Natl Acad Sci 117(36):22303–22310
109. Vignal A, Milan D, SanCristobal M, Eggen A (2002) Genet Sel Evol 34:275
110. Vincent AC, Sadovy de Mitcheson YJ, Fowler SL, Lieberman S (2014) The role of CITES in the conservation of marine fishes subject to international trade. Fish Fish 15(4):563–592
111. Wheat CW (2010) Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. Genetica 138(4):433–451
112. Wheat RE, Allen JM, Miller SDL, Wilmers CC, Levi T (2016) Environmental DNA from residual saliva for efficient noninvasive genetic monitoring of Brown Bears (*Ursus arctos*). PLoS ONE 11:e0165259
113. Winker K, Glenn TC, Faircloth BC (2018) Ultraconserved elements (UCEs) illuminate the population genomics of a recent, high-latitude avian speciation event. PeerJ 6:e5735
114. Woodall LC, Jones R, Zimmerman B, Guillaume S, Stubbington T, Shaw P, Koldewey HJ (2012) Partial fin-clipping as an effective tool for tissue sampling seahorses, *Hippocampus* spp. J Mar Biol Assoc UK 92(6):1427–1432
115. Worley KC, Gibbs RA (2010) Decoding a national treasure: the giant-panda genome is the first reported de novo assembly of a large mammalian genome achieved using next-generation sequencing methods. The feat reflects a trend towards ever-decreasing genome-sequencing costs. Nature 463(7279):303–305
116. Zhao Q (2018) In: Proceedings of the 2018 5th international conference on bioinformatics research and applications, Association for Computing Machinery, New York, NY, USA, ICBRA'18, pp. 8–15. https://doi.org/10.1145/3309129.3309134
117. Zoonomia Consortium (2020) A comparative genomics multitool for scientific discovery and conservation. Nature 587(7833):240
118. Zulkefli NS, Kim K-H, Hwang S-J (2019) Effects of microbial activity and environmental parameters on the degradation of extracellular environmental DNA from a Eutrophic Lake. Int J Environ Resh Public Health 16:3339

Anubhab Khan is a wildlife genomics expert. He has been researching genetics of small isolated populations for past several years and has created and analyzed large scale genome sequencing data for tigers, elephants and small cats among others. He is keen about population genetics, wildlife conservation and genome sequencing technologies. He is passionate about finding solutions to technology disparity in the world by either making advanced technologies and expertise available or by developing techniques that are affordable and accessible to all.

Abhinav Tyagi is a molecular ecologist and PhD scholar at National Centre for Biological Sciences, Bangalore. He is investigating anthropogenic impacts on the connectivity of multiple species of herbivores and carnivores in central India. He is interested in understanding species-specific responses towards habitat fragmentation and also keen on developing novel methods for conservation genomics that enables the use of non-invasive samples for wildlife research.