Intraspecific Diversity of Congeneric Species Along an Elevational Gradient

विद्या वाचस्पति की
उपाधि की अपेक्षाओं की आंशिक पूर्ति में प्रस्तुत शोध प्रबंध

A thesis submitted in partial fulfilment of the requirements of the degree of
Doctor of Philosophy

द्वारा / By

Debjyoti Dutta

देबज्योति दत्ता

पंजीकरण सं. / Registration No.:
20153397

ध प्रबंध पर्यवेक्षक / Thesis Supervisor:

Ramana Athreya

रमाना अथरेया

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE

2025

To Ma, Baba, Didi

Thank You for Everything

CERTIFICATE

Certified that the work incorporated in the thesis entitled **Intraspecific Diversity of Congeneric Species along an Elevational Gradient** Submitted by **Debjyoti Dutta** was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

Ramana Athreya

(Supervisor)

Date: 18 Dec 2025

Declaration by Student

Name of Student: Debjyoti Dutta

Registration No.: 20153397

Thesis Supervisor: Ramana Athreya

Department: Biology

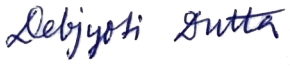
Date of joining program: 31/05/2015

Date of Pre-Synopsis Seminar: 20/06/2025

Title of Thesis: **Intraspecific Diversity of Congeneric Species along an Elevational Gradient**

I declare that this written submission represents my idea in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I declare that I have acknowledged collaborative work and discussions wherever such work has been included. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Thus work reported in this thesis s the original work done by me under the guidance of Dr. Ramana Athreya



Debjyoti Dutta

Date: 19/12/2025

Acknowledgements

I always wanted to get trained in acquiring skills for identifying and solving problems which is essential for any kind of research. The very progress of any individual is based upon rational thinking and developing a scientific temper. Doing a PhD became a natural and obvious choice. It provides a learning curve and as long as a person learns some skill, gains some wisdom that will be valuable in the future, it is worth it. To that end, I find myself to have acquired skills in experimental design, molecular techniques and in coding.

I must thank my supervisor Dr. Ramana Athreya for his guidance and indeed I get the feeling that if there is anything practically useful, I have learnt that during my PhD. He has always pushed me to be more organised (something which still needs enormous improvement from my end) to generate quality data and overall be a better scientist. One key lesson I learnt from him is that no task is ever so simple that it does not require thoughtful engagement. I thank him for investing his valuable time on me.

I thank my RAC members Dr. Deepak Barua and Dr. Suvendra Ray for providing valuable inputs and words of encouragement. I thank IISER Pune for funding my Phd.

I joined Ramana's lab because I wanted to have an understanding of techniques in molecular biology. I had never worked in a wet lab before and it was Mansi who both helped and guided me during my initial days and for any skills that I have picked up later, I have her to thank. I also thank Alakananda for clearing doubts about my scripts and for helping me out in installing the sequencing software. I thank Anurag Mishra and Rohan Pandit, past members of our lab for sampling of the frogs and birds and it is their sampling efforts which made this thesis possible. Although I have never been to field, I do thank all the people stationed there whose hard work I know is involved in answering my research questions.

I express my heartfelt gratitude to current lab members Netra, Sonali, Divya, Prathamesh and Unnikrishnan who made lab life fun and rescued me from gloom. I have cherished interacting with them, helping them out and in turn getting help from them. I owe you guys a treat.

Our lab has a collaboration with GenePath Dx, a molecular diagnostic company and it is my privilege to have worked in their lab. They were kind enough to allow me to use their facility for sequencing. I especially thank Dr. Nikhil, Shatakshi, Kunal for providing me exposure advanced molecular techniques.

I thank my parents and my sister for being patient with me and never losing faith in me. They always backed me through tough times, helped me out when I was doubting myself and never allowed any pressure of family to come on to me.

Finally I must thank all the pups of IISER who always gleefully run towards me and are always happy see me whether I have treats or not. They have been little bundles of joy.

Synopsis

The pattern of decreasing diversity with increasing latitude was one of the earliest patterns noted by ecologists. This pattern has been ascribed to the differing environmental conditions across the latitudinal span. Elevational gradients also comprise divergent environmental conditions but over much smaller geographic distances. For example, the temperature lapse rate of 0.65°C per 100 m results in a change of ~5°C for every 750 m of elevation which is the same difference as across 9 degrees of latitude translating to a geographical distance of 1000 km. Montane populations of species occur in close proximity to environments (at other elevations) to which they are not optimally adapted. Their persistence in such conditions, at the edge of the range, indicates their considerable ability to adapt. Local adaptation results in genetic differentiation and is in turn affected by intraspecific genetic diversity. Higher the intraspecific diversity, greater the probability that a population would have alleles that enhance fitness in a local environment which would then increase in frequency for that population. This may or may not lead to species level differentiation contributing to species diversity.

Differential range shifts amongst montane species due to climate change may bring previously isolated species pairs into contact with each other (e.g. at a ridge). This can cause a change in diversity profiles as well as in interspecific competition dynamics in a community. Therefore, elevational gradients are an important ecological setting to study the dynamics of variation within species; the small distance between populations reduces the influence of distance – the other driver of diversity.

This thesis investigated the patterns of intraspecific diversity and genetic differentiation of frogs of the *Raorchestes* genus (family: Rhacophoridae) along a compact but steep elevational transect in the eastern Himalayas. After finding sharp transitions between species ranges in 3 contiguous frog species, we investigated the degree of overlap of adjacent congeneric species pairs in the much larger bird data set (comprising 245 species) along the same elevational transect.

The investigation of intraspecific diversity and genetic differentiation along elevation requires large amount of DNA sequence data to reliably discern the small quantum of expected difference. In the absence of knowledge of the specific genetic loci which are implicated in elevational adaptation, one can only try to measure population differentiation due to (partial) isolation using neutral

markers. We know that cycles of climate change have a periodicity of ~100 kyr. The most recent major event of climatic change was the end of the ice age ~15 kyr ago. An average mutation rate for frogs of ~0.05 per site per Myr, translates to 5 mutations per 10 kb per 10 kyr. The poisson noise on the number of mutations will be ~2.2. Therefore, DNA fragments of length 10 to 20 kb are required to sufficiently sample mutations above the poisson noise which can resolve divergences among populations that differentiated ~15 kyr ago.

Therefore, a major part of this thesis consists of the development of a molecular technique which can provide sufficient DNA data for a large number of samples at a low price – 10-50 kb per sample at a price of 5-10 USD. As part of this – as required by the Nanopore NGS platform needed for the procedure – we also developed a Bayesian technique to estimate the error on each consensus base in order to limit the data to loci where the error was less than 1 in 10^3.

With these techniques, we successfully sequenced ~14 kb of DNA for 172 individuals across 500-2400 m elevational transect in Eaglenest Wildlife Sanctuary in Arunchal Pradesh, India. A phylogenetic tree showed that the samples segregated into 3 elevationally separated clades. The average pairwise sequence divergence of the clades was ~7%, well above the fiducial 5% divergence seen in most frogs. Thus our study involved 3 species, which we label by their localities: Khellong species at low elevation (550-1100 m; n = 48), Sessni species at mid elevation (750-1750 m; n = 31) and Bompu species at high elevation (1650-2350; n = 93). We found the mid-elevation Sessni species to have the most intraspecific genetic diversity. It is possible that the Sessni species is shielded from range edge decimation to which lower Khellong and higher Bompu may be periodically exposed during climate cycle extremities. During global maxima, the increase in temperature would push Bompu species to the edge of the available elevation at the ridge at 3250 m elevation. Any further warming may lead to population and diversity loss. The same may happen for the Khellong species during global minima, with Khellong individuals, from the foothills, being unable to adapt to the flood plain forest and undergoing periodic loss of intraspecific diversity. Sessni, being in the middle, is protected from the effects of such range shifts and can retain its diversity.

We also tested for signature of local adaptation by measuring genetic differentiation with elevational separation within a species. We detected a significant relationship between genetic distance and geographic distance only for the Bompu species. This pattern may arise because higher elevation species are known to experience more stressful conditions (e.g. stress dominance hypothesis). The absence of any physical barriers strongly suggests that the differentiation is the

outcome of local adaptation. The genetic differentiation for Bompu was detected across a geographical distance of just 2.8 km which is much lower than what is reported in any other study – intraspecific differentiation is typically detected over distances of many tens to hundreds of kilometres.

We found that the width of the transition zone for both Khellong-Sessni and Sessni-Bompu was only ~100 m. There is much debate on whether range edges are determined by competition or climatic constraints. We could not answer this with just the data from frogs. Therefore, we used a large data set of ~16000 records of 245 birds species to determine the distribution of the width of range overlap between congeneric species. This width was estimated by fitting a logistic curve to the abundance data spanning the zone of overlap.

We analysed the range overlap of 107 congeneric species pairs, which had been recorded in sufficient abundance and also replaced each other elevationally. A large fraction of species pairs had a transition width of less than 100 m. However, a curated dataset of 69 pairs did not show any correlation between transition width on one hand and phylogenetic distance on the other.

In summary, this thesis introduces a novel strategy for generating several tens of thousands of base pair data of DNA at a much lower price than either Sanger or standard NGS techniques. This is very useful for all studies which need to measure intraspecific diversity. Analysis of intraspecific genetic diversity indicated that intraspecific genetic diversity was highest for the mid-elevation species; the lower diversity in the higher and the lower elevation species may be related to decimation of populations due to climate cycling. Investigation of genetic differentiation by analysing relationship of genetic distance with elevational distance detected a positive correlation only for the high elevation Bompu species. Finally, the sharp transition in ranges of the 2 pairs of frog species was investigated in greater detail using 107 pairs of congeneric bird species. We found that a significant number of species pairs do have sharp range transitions, which suggests that they are determined by competition.

Contents

Chapter 1: Introduction 14

Chapter 2. Methods of Field and Lab 20

2.1 ***Raorchestes*** **Frogs along an Elevational Gradient** 20
Study Region – Arunachal Pradesh in the Eastern Himalayas
Eaglenest Wildlife Sanctuary
Field Sampling of *Raorchestes frogs*

2.2 **Intraspecific Population Structure using Molecular Methods** 23
DNA Extraction from Tissue Samples of Individual Frogs
PCR selection and amplification of target fragments
Genome Target Selection and Primer Design
PCR amplification
Index PCR
PCR Product Purification
DNA Quantification
Next Generation Sequencing (NGS)

Chapter 3. A Consensus Base-Calling Procedure for NGS Data 46

3.1 **Introduction** 46

3.2 **Base-calling Algorithm** 49
Reference-Set
Q scores
Fractional Distribution of Read-bases at a Locus
Reference-to-Read probability
Calling the Consensus-Base

3.3 **Discussion** 60

Chapter 4. Intraspecific Diversity Along Elevational Gradients 62

4.1 **Introduction** 62
Biodiversity
Elevational influence on genetic diversity
This Work

4.2 **Data and Methods** 67
Basic Data
Pairwise Phylogenetic Distance

4.3 **Results** 69
Variation in intraspecific diversity with elevation
Phylogenetic distance versus Elevational distance within spec**ies**

4.4 **Discussion** 72

Chapter 5. Range Overlap in Congeneric Species Pairs 75

5.1 Introduction 75

5.2 Data and Methods 78

Raorchestes Frogs

Bird data

Estimation of Transition Width

Correlation between Overlap Width and Functional Traits

5.3 Results 81

Raorchestes Frogs

Bird Community

5.4 Discussion 86

Conclusion 90

References 91

Chapter1

Introduction

Some regions on earth tend to harbour higher species diversity and abundance compared to others. Even populations of the same species are not uniformly distributed across its habitable space. This phenomenon of uneven biodiversity at large scales is well documented (Wallace 1876, Currie 1991, McGill 2010, Wiens & Donoghue 2004, Gizachew, 2021). One of the most well-known patterns depicted in ecology is the Latitudinal Diversity Gradient which shows higher species richness towards the equator that gradually decline with increasing latitude (Pianka 1966). Several mechanisms have been proposed as explanations for latitudinal gradients of diversity such as: (i) Evolutionary Speed Hypothesis (Rohde 1992, Allen & Gillooly 2006, Mittelbach et al 2007) which claims that warmer temperatures drive increase in metabolic and mutation rates and shortens generation time leading to faster speciation in the tropics, (ii) Higher speciation events in the tropics because the tropical regions are older and more stable that provide more time to diversify for the lineages that originated over there (Pianka 1966, Wiens & Donoghue 2004, Kerkhoff et al 2014), (iii) Energy/productivity hypothesis (Currie 1991, Hawkins et al 2003, Fritz et al 2016) postulating that tropical regions are recipients of higher solar energy, leading to higher primary productivity, thereby supporting more number of individuals belonging to different species resulting in higher diversity, (iv) Higher speciation rate in the tropics due to stronger biotic interactions and environmental heterogeneity (Dobzhansky 1950, Ricklefs 2006), (v) Lower extinction rates in the tropics relative to temperate regions because temperate regions experience glaciations and climatic fluctuations which leads to local extinction and loss of diversity (MacArthur 1984, Mittelbach et al 2007), (vi) Tropical regions have historically covered large areas that support more species and increases speciation opportunities (Rosenzweig & Sandlin 1997, Schoener 2010, Marin et al 2018), and (vii) Biotic interactions hypothesis which postulates that strength of biotic interactions like competition, predation and mutualism promotes specialisation resulting in niche partitioning and ultimately driving speciation (Pianka 1966, Schemske et al 2009)

Across the large geographical span of latitudinal gradients the environment co-varies with several other factors like evolutionary history, area, species pools and productivity. This makes it difficult to identify the influence of environmental factors on species diversity. Elevational gradients, especially compact ones, offer a context in which a wide range of environmental variation occurs within the same regional species pool keeping biogeographical history relatively constant. This enables elevational gradients to be excellent natural laboratories to specifically test for the influence of divergent environments on diversity patterns. They also are useful to predict response of species to climatic changes in future.

Elevational transects host steeper gradients in abiotic conditions than latitudinal transects. For example, the lapse rate for temperature is 1°C for every 111 km along latitude but only 165 m for elevation, which can correspond to as little as 1 km of ground distance along slopes which are steep but still accessible. Such short distances makes elevational studies logistically easier in some respects.

Intraspecific genetic diversity and differentiation along elevational gradients

The role of elevational gradients as natural laboratories of evolution has been extensively studied to determine the mechanisms by which variations accumulate within elevational populations. Numerous studies across taxa demonstrate that trait variation, genetic diversity and differentiation are structured along elevational gradients. Signatures of elevational clines of functional traits among multiple taxa serve as the most distinct evidence of elevationally induced intraspecific variation, e.g. leaf thickness and area for plants (Umaña & Swenson 2019), increase in body size with elevation for insects (Sharma et al. 2023), migration and breeding trait variation along elevation for birds (Lundblad & Conway 2020) and variation in skull morphology of mammals (Feijo et al. 2019).

In addition to morphological variation, molecular data have also uncovered genetic variation. Decline in genetic diversity along elevation has been reported for plants as higher elevation populations are expected to be more isolated (Zormpa et al 2025). The same pattern has also been reported for Montane mayflies (Polato et al. 2017) and long-toed salamander (Giordano et al. 2007). Higher elevations harbour stressful climatic conditions due to which species develop adaptations specific to such environments. This reduces gene flow to such environments as introduction of

maladapted alleles reduces their fitness and filters out alleles arising due to mutation as well. As a result diversity loss along elevation becomes an expected pattern.

Elevational gradients also exhibit patterns of genetic differentiation as observed for plants where differentiation increases especially towards the range edge (Daco et al. 2022), distinct gene pools across elevation are observed in alpine plants (Felkel et al. 2023) and in amphibians (*Bufo bufo*) strong signal of elevational differentiation for one metric of differentiation over another (QST>FST) (Luquet et al. 2015). Since montane environments are not spatially isolated by large distance, genetic differentiation in this context points to isolation by environment more than isolation by distance, thereby indicating the influence of selection pressures in generating intraspecific variation.

In this thesis, we investigated the patterns of intraspecific diversity and differentiation along a steep elevational transect in 3 species of *Raorchestes frogs. The elevational ranges of the species span* only a few kilometres of ground distance, which is similar to the dispersal distance observed elsewhere (e.g. Semlitsch, 2000; Funk et al, 2005). As a consequence, one does not expect any genetic structure across the range unless there is local (elevational) adaptation, with temperature and humidity being the most likely factors influencing the same (Lillywhite & Navas 2006, Spranger et al. 2024, Dhib et al. 2022, Campbell et al. 2024, Gass et al. 2024). Nevertheless, intraspecific genetic variation is expected to be very small, occurring in 0.01-0.1 % of the bases.

We had to solve two issues arising from the expected weak genetic differentiation signal: (i) estimating the error of base-calling at a sequenced locus to avoid contaminating good data by the bad – we did not come across any such algorithm in published literature, and (ii) reducing the cost to generate sufficient sequence data for a large number of samples.

Genetic variation within a species was first estimated during the 1960s using protein electrophoresis to detect variation in enzymes (allozymes) among individuals. Restriction Fragment Length Polymorphism (RFLP) was the first technique developed to use DNA as a molecular marker to detect genetic variation. It compared the lengths of DNA fragments among individuals after restriction enzyme digestion. While it was more informative than proteins, it could not differentiate between individuals in which fragment sizes remained the same but their actual sequences differed. The invention of Polymerase Chain Reaction (PCR; Mullis & Faloona, 1987) and Sanger

sequencing (Sanger et al. 1977) made it possible to amplify and sequence specific fragments of DNA which could then be compared between individuals.

The Sanger technique is only capable of sequencing DNA fragments of ~1000 bases. Therefore, one needs to sequence a large number of fragments for statistically significant detection of intraspecific genetic variation. This can be both expensive and time and labour intensive. Furthermore, PCR amplification of short fragments (required for Sanger sequencing) in the intronic regions is a difficult proposition. This is a problem since the highly variable intronic sections are likely to provide the strongest signal of genetic variation.

The beginning of the mid-2000s saw a dramatic shift in sequencing technology, going from sequencing one DNA fragment at a time to massive parallel sequencing of millions of DNA fragments simultaneously. This group of (2^{nd} generation) sequencing technologies came to be known as Next-Generation Sequencing (NGS; e.g. Roche 454, Illumina Inc). NGS dramatically improved speed and lowered cost per sequence base, and facilitated the sequencing of whole genomes. NGS generates sequence data by fragmenting the input DNA, sequencing those fragments (typically 50-300 bases long) and finally assembling those fragmented reads back to the original molecule using established bioinformatic pipelines. The error rate of NGS sequenced DNA is quite low at ~1% but it suffers from uneven representation of different segments across the sequenced molecule during library preparation and difficulty in resolving long repeat regions within the genome.

The latest iteration of NGS – the third-generation (e.g. PacBio, Oxford nanopore) – can sequence intact DNA fragments upto tens of thousands of bases in length.This avoids the problem associated with assembly of repeat regions and also minimises the uneven coverage mentioned previously.

While NGS has indeed dramatically reduced the cost per sequenced base (compared to Sanger), it is still very expensive to sequence several hundred specimens. In this project we developed a novel strategy which combined the selectivity of PCR to amplify long fragments across introns and sequenced them using Oxford Nanopore to obtain sufficient data to securely detect intraspecific variation in our target species.

Overlap zones in species ranges

The geographical range of a species is determined by a combination of factors including environmental gradient, genetic and phenotypic variation, adaptability, vagility and interspecific competition (Kirkpatrick & Barton 1997). The steep environmental gradients along elevational transects compresses the distance between the optimal and suboptimal zones for a locally (elevationally) adapted population of a species. Therefore, constraints of vagility, which can be considerable along latitudinal gradients, are less likely to play a role in the mountains (Janzen 1967, McCain & Grytnes 2010). Therefore, there is a high probability that higher fecundity in the optimal zone will lead to considerable overflow into the non-optimal zones. The ability of this overflow (population) to establish themselves in the suboptimal zone will depend on their adaptability and relative fitness vis-a-vis competitors.

Along a mountain slope, the short spatial distance and high species diversity is expected to result in a higher probability of contact between species adapted to different environments (Körner 2007, Colwell et al. 2008). This is especially interesting in the context of climate change where species ranges are expected to shift in elevation while tracking their optimal niche (Parmesan 2006, Chen et al. 2011). This can bring previously separated species into contact or increase their extent of overlap (Lenoir et al. 2008, Jackson & Overpeck 2000), increasing the potential for competitive interactions. The competition is expected to be more intense in the case of congeners since they are likely to share similar functional traits (Tilman 1982, Losos 2008, Kraft et al. 2015). Therefore, extent of overlap between congeners can be a measure of the relative influence of competition and environmental factors in setting range limits.

The structure of this thesis is as follows:

In Chapter 2, we introduce the study taxon, the area in which it was sampled and the wet-lab procedures required for generating the molecular sequence data.

In Chapter 3, we present a novel Bayesian approach to base-calling at each locus of the consensus of NGS reads for each specimen. The importance of this technique is that it provides an estimate of the error probability of each individual consensus base call. This offers the use of a threshold to identify the less reliable base-calls in the analysis.

In Chapter 4, we present the pattern of genetic diversity and differentiation of *Raorchestes frog* populations along an elevational gradient. We first show that our cryptically coloured frogs are actually three elevationally segregated species. We then show that even these closely related species differ in the degree of intraspecific diversity, which may be related to location of their range and climate. We also show evidence for genetic differentiation with elevation in one species, which is likely due to local adaptation.

In Chapter 5, we investigated the nature and extent of range overlap of our frog species pairs and compared it to congeneric bird pairs whose data was also available from the same region. Closely related congeneric pairs share ecological traits making them prone to competition which can result in mutual restriction of the species' ranges; the observed range limits need not be a reflection of maladaptation to abiotic conditions. A comparison between frogs and birds would therefore elucidate whether a particular range overlap pattern is a general feature of congeneric species inhabiting along an elevational gradient.

Chapter 2

Methods of Field and Lab

Section 2.1: *Raorchestes Frogs along an Elevational Gradient*

Section 2.1.1: Study Region – Arunachal Pradesh in the Eastern Himalayas

The eastern Himalayas stretch over 800 km covering eastern Nepal, Sikkim (India), Bhutan, Arunachal Pradesh (India), northwest Yunnan (China), southeast Tibet (China) and northern Myanmar. Covering a total area of 524,190 $km^{2,}$, the Eastern Himalayas span a complex and rugged topographic landscape (Zomer et al. 2002) and are bounded by the Brahmaputra valley to their south. This region was earlier considered to lie at the interface of the Indo-Malayan and Palaearctic biogeographic realms (Wallace, 1876) but recent revision in biogeography based on phylogenetic relationships puts it at the junction of the Oriental and Sino-Japanese biogeographic regions (Holt et al. 2013).

The eastern Himalayas are among the four global hotspots of biodiversity listed for India as per the Myers-Mittermeier framework (Myers et al. 2000). The species richness of breeding birds is second only to Andes-Amazonia and at par with the Rift Valley in Africa (Orme et al., 2005). It has the highest diversity of oscine passerines in the world (Price et al., 2014). Yet apart from birds, its diversity is only beginning to be explored in recent years (Athreya, 2006). WWF-India recently reported discovery of 211 new species between 2009 and 2014 from this region.

Arunachal Pradesh in north-east India covers the easternmost part of the Himalayas. Nearly 70% of its area is covered with mountains and 82% is under forest cover (Paul 2005). It constitutes less than 3% of India's land area but hosts almost two-thirds of its biodiversity.

This region encompasses a steep elevational gradient ranging from 100 m on its southern slope to ~7000 m at its border with Tibet. The rainfall varies from 1500 mm along northern slopes to over 3500 mm in the southern slope (Choudhury 2003). The heavy rainfall is drained by deep north-south river gorges which become the tributaries of Brahmaputra making it the largest river basin in

the Indian subcontinent. The vast elevational gradient and variation in precipitation has aptly earned this region the 6th rank among the hottest of biodiversity hotspots (Myers et al. 2000) and it also features among the top 200 eco-regions of the world (Olson and Dinerstein, 1998).

Section 2.1.2: Eaglenest Wildlife Sanctuary

Our study was carried out in Eaglenest Wildlife Sanctuary (EWS) in the West Kameng District in the western part of Arunachal Pradesh. It is a small protected area of 218 km^2. It hosts pristine forests across an elevational range of 100 to 3250 m. Vehicular access along a dirt track traversing most of the elevational range greatly facilitates research in the area (Athreya, 2006).

In EWS, elevations below 500 m experience warm temperatures during the months of May to June (>25 °C), regions between 500 to 1900 m have temperature ranging from 10 to 18 °C. February is the coldest month of the year with occasional snowfall above 2000 m. It experiences major amount of rainfall during the months of June to October. Annual rainfall reportedly varies from less than 1500 mm on the northern slopes to greater than 3000 mm on the southern slopes.

The large elevational range in combination with extensive rainfall harbours diverse habitat types in Eaglenest ranging from tropical wet evergreen (below 900 m) to coniferous temperate forests (above 2800 m). The area was almost entirely unexplored prior to 1995 before the advent of the Eaglenest Biodiversity Project in 2003 (Athreya, 2006; Agarwal et al. 2010; Mungee & Athreya, 2021) and others (Srinivasan et al. 2010; Marathe et al 2020). Currently, about 425 bird species, over 1500 lepidoptera species, 75 herpetofauna and 40 species of mammals have been recorded from this region underscoring its importance as an excellent place for ecological research (Athreya 2006).

Section 2.1.3: Field Sampling of *Raorchestes* frogs

Raorchestes are a genus of bush frogs that belong to the family Rhacophoridae which are found in South and South-East Asia. These frogs occur in the Western Ghats and in the north-eastern states of India. It is a speciose genus that displays a high degree of endemism, comprising almost one-sixth extant frog species in India (Dinesh et al. 2015). Their high abundance (Biju & Bossuyt 2003; personal observation in EWS) – hence negligible conservation concerns – and easy detection due to

their distinctive call (at the generic but not species, level) make them an ideal candidate for sampling of large number of individuals.

Raorchestes individuals were sampled for location and toe tissue (research permit from the Govt of Arunachal Pradesh No. CWL/D/21(106)/2012/677-89) by a team of undergraduate students and project staff under the supervision of Anurag Mishra for his undergraduate project (Mishra A. 2015). They were sampled during May-June 2012 along the southern slope of Eaglenest Wildlife Sanctuary between the elevations of 500 m and 2400 m. Intensive sampling efforts beyond 2400 m yielded very few individuals, while the areas below 500 m could not be sampled due to logistical issues. Since, *Raorchestes* is a cryptic species group, an estimate of the species' elevational range was not available at the time of sampling. Previous sampling in the area indicated that other Rhacophorid genera had 3-5 species (Athreya, 2006) along that elevational gradient, suggesting an average species elevational range of about 600 m. Therefore, the team sampled 10-15 individuals per elevational band of 100 m. The individuals (all males?) were tracked by their calls. Some of them were euthanised using MS-222 while tissue for others were collected via toe clips. Tissue samples for molecular analysis were preserved in 99% ethanol.

Figure 2.1: Images of representative *Raorchestes* individuals sampled in Eaglenest. Photo credit: Anurag Mishra

Section 2.2: Intraspecific Population Structure using Molecular Methods

The investigation of intraspecific diversity and genetic differentiation along elevation requires large amount of DNA sequence data to reliably discern the small quantum of expected difference. In the absence of knowledge of the specific genetic loci which are implicated in elevational adaptation, one can only try to measure population differentiation due to (partial) isolation using neutral markers. We know that cycles of climate change have a periodicity of ~100 kyr. The most recent major event of climatic change was the end of the ice age ~15 kyr ago. An average mutation rate for frogs of ~0.05 per site per Myr (Crawford 2003, Lau et al. 2020) translates to 5 mutations per 10 kb per 10 kyr. The poisson noise on the number of mutations will be ~2.2. Therefore, DNA fragments of length 10 to 40 kb are required to sufficiently sample mutations above the poisson noise which can resolve divergences among populations that differentiated ~15 kyr ago.

The vast majority of studies, until recently, in molecular ecology and phylogenetics have relied on sequencing relatively small fragments of mitochondrial DNA in the range of 1000-1500 bp (Avise et al. 1987, Hebert et al. 2003, Galtier et al. 2009). This limit is imposed by Sanger sequencing (Sanger et al. 1977, Kocher et al. 1989, Shendure and Ji. 2008, Ekblom and Galindo. 2011) which cannot sequence DNA fragments much longer than 1000 bp. Therefore, one has to build up sufficient DNA fragment length (per specimen) by PCR amplification and sequencing of many short fragments. Also Sanger sequencing deals with single samples. This makes the procedure either labour and time intensive or expensive or both, especially for intraspecific studies wherein the time to the last common ancestor and hence intraspecific genetic divergence, is much smaller.

The use of nuclear markers (with the size limit imposed by Sanger sequencing) are less useful since the exons tend to be more conserved than mitochondrial DNA (Hwang and Kim 1999) and short fragments of the more variable introns are difficult to amplify using PCR.

Next Generation Sequencing (NGS) platforms (Margulies et al. 2005, Shendure and Ji. 2008, Metzker 2010) offer the advantage of parallel sequencing of multiple samples in a single run, thus, enormously reducing time and cost per sample. However, the initial NGS procedures were designed to sequence the entire genome. While the cost per base is indeed very low for sequencing the entire genome, the cost per specimen is still very high (upwards of INR 200,000/USD 250 for large genomes like that of frogs). It is particularly a waste when the project does not even require that

large amount of data – the cost per "necessary" data would be even higher. The technique of RADseq (Restriction Site Associated DNA Sequencing; Baird et al. 2008, Davey et al. 2011) enormously reduces sequencing costs but still costs INR 15-20,000 (USD 175-200) per sample. Furthermore, the effectively random nature of the location of restriction sites, on which the technique is based, precludes the possibility of selecting particular loci.

We developed a strategy that combines selectivity and library prep of individual samples using long-range PCR, and can multiplex many hundreds (or even thousands) of samples to make use of inexpensive sequencing using NGS. This process can generate 10-40 kb DNA sequences at less than INR 1000 (USD 10) per sample – making it economical to study intraspecific variation of hundreds of specimens.

The principal steps in our strategy included

1. DNA extraction from Tissue samples of individual frogs
2. Genome Target Selection and Primer Design
3. Amplification of target region and attachment of Indexes for sequencing using PCR
4. Standardizing protocol for multiplexed sequencing via NGS

Section 2.2.1: DNA Extraction from Tissue Samples

DNA was extracted from frog tissue using Invitrogen's ChargeSwitch gDNA mini-tissue kit (CS11204). The procedure makes use of buffers of different pH to bind the beads to DNA or release them from the surface of paramagnetic beads. The bound DNA is washed by acidic solutions to remove the impurities and then eluted using neutral or basic solutions. Unlike silica columns which yield DNA fragments of ~3 kb size, ChargeSwitch beads yield DNA fragments well in excess of 10 kb which is essential for amplifying large DNA fragments.

The product recommends 10-25 mg of (mouse) tissue for the volume of the (25) standard reaction volumes in the kit. However, our samples consisted of small pieces of thigh muscle or toe clips amounting to no more than a few milligrams. We ran several assays to optimise DNA extraction using only a tenth of the prescribed ChargeSwitch bead volume. The rest of the protocol was as per the manufacturer's recommendation. The purified DNA sticking to the beads was eluted into two vials of 30 μL of 1X-TE. The extracted genomic DNA were run on 1% agarose gel to check for size

and integrity. Their concentration was measured using a NanoDrop (Desjardins and Conklin. 2010). Second elutions had slightly lower concentrations and less distinct band as expected. The amount of genomic DNA obtained ranged from 0.2 to 2.5 µg. Their size was likely to be in excess of 10 kb.

Section. 2.2.2: PCR selection and amplification of target fragments

We aimed to sequence two DNA fragments of size 5-10 kb, one each from mitochondrial genome and the nuclear genome. This was deemed sufficient for detecting genetic variation in order to determine differences between populations of a species. We targeted loci associated with the standard genes which had been used in previous phylogenetic studies (Table 2.1).

Section 2.2.3: Genome Target Selection and Primer Design

Mitochondrial fragment

The entire mitochondrial genome ranges approximately from 15 kb to 21 kb and is much smaller than the nuclear genome. Mitochondrial whole genome sequences are available for a wide range of anuran taxa which facilitates the design of primers in conserved regions. We decided to sequence a large single fragment of mitochondrial DNA of ~7 kb which included rRNA, tRNA and several protein coding genes. This fragment included the gene COX1, which is well known for its use as a molecular barcode for species delineation (Hebert et al. 2005). A survey of published literature revealed that 12S, 16S and tRNA val were popular regions for locating primers (Table 2.1). The forward primer was designed in this region. We could not find a published primer downstream of COX1 and hence designed a new primer for our taxa.

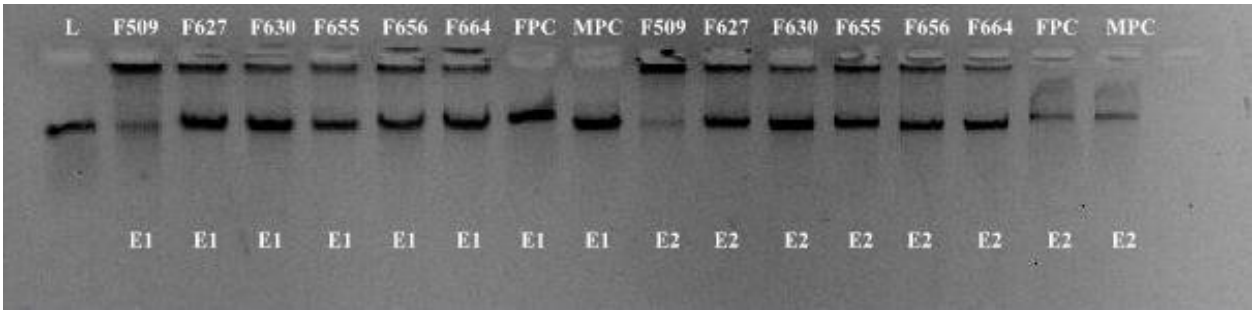


Figure 2.2: Electrophoresis of genomic DNA extracted using the ChargeSwitch protocol. The ladder (L) is a 7000 bp fragment. FPC and MPC are frog and mouse tissue positive controls. The rest of the lanes are for the study samples including the first (E1) and second (E2) elutes of DNA. The extracted DNA was much larger than the 7000 bp marker.

We downloaded mtDNA sequences of 5 species from Rhacophoridae family and aligned the sequence between 12S rRNA and tRNA Lys. The sequences belong to *Rhacophorus schlegelii* (Genbank ID: NC_007178), *Polypedates megacephalus* (Genbank ID: AY458598), *Polypedates braueri* (Genbank ID: NC_042797), *Rhacophorus dennysi* (Genbank ID: NC_027452), Buergeria buergeri (Genbank ID: AB127977). We designed two pairs of forward and reverse primers, with one pair meant as a backup in case the first set did not work.

The internal primers served the purpose of redundancy, in case the combination of two external primers failed to amplify. The locations of these primers on the mtDNA alignment are shown in Figures 2.3-2.5.

Table 2.1: Mitochondrial gene loci of published primers

Gene	Primer paper
12S	Darst and Cannatella, 2004
12S	Faivovich et al, 2004
12S	Titus and Larson, 1996
12S	Wilkinson et al, 1996
12S	Graybeal, 1997
12S	Goebel et al, 1999
16S	Darst and Cannatella, 2004
16S	Faivovich et al, 2004
16S	Palumbi et al, 1991
16S	Hedges, 1994
16S	Titus and Larson, 1996
16S	Wilkinson et al, 1996
tRNA Val	Darst and Cannatella, 2003

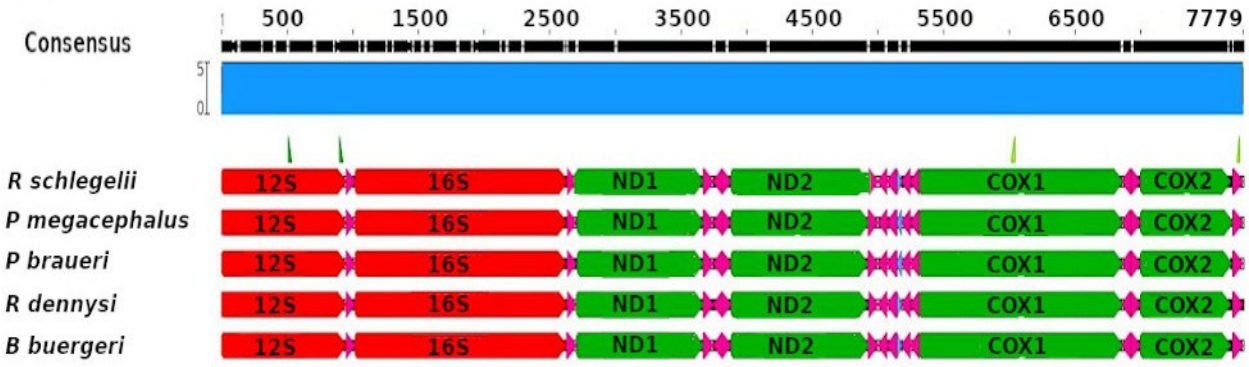


Figure 2.3: Alignment of partial mitochondrial DNA sequence of 5 species from Rhacophoridae family. Deep green triangles above the alignment are forward primers and light green triangles are reverse primers. All the genes are annotated. tRNA labels (pink triangles) are too small to be displayed.

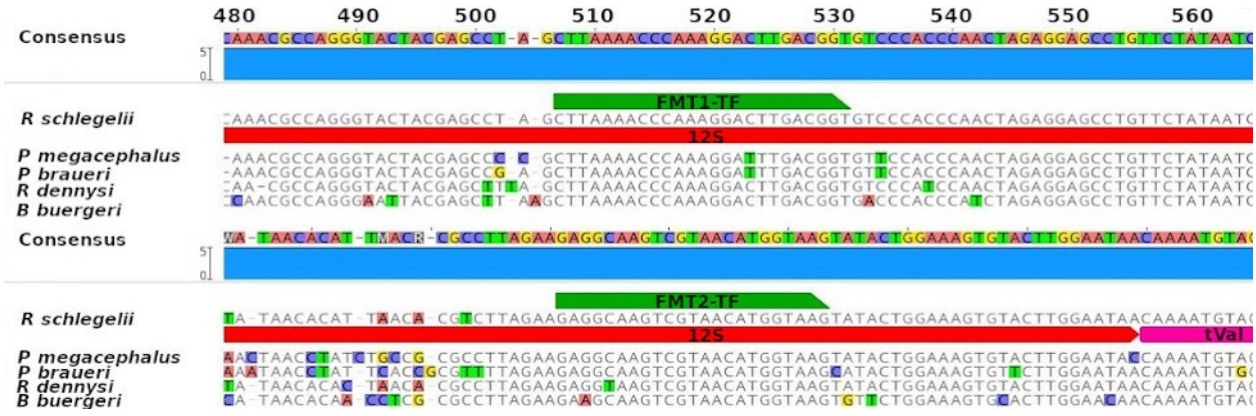


Figure 2.4: Alignment of sequences from 5 Rhacophoridae species showing the regions where the forward primers (FMT1_TF and FMT2_TF) were located.

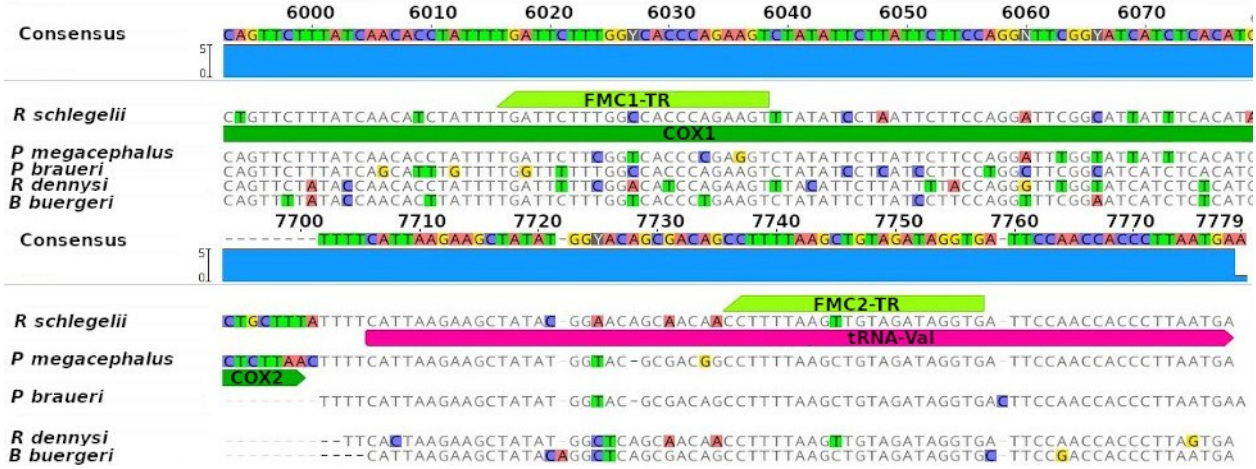


Figure 2.5: Alignment of sequences from 5 Rhacophoridae species showing the regions where the reverse primers (FMC1_TR and FMC2_TR) were located.

Table 2.2: Candidate nuclear genes considered for the sequencing campaign.

Gene	Number of Exons	UTR	Primer location	Primer paper
Rhodopsin	5	---	1st & 4th exon	Bossuyt et al, 2000
Tyrosinase	5	---	1st exon	Bossuyt et al, 2000
CXCR4	2	1st exon	2nd exon	Biju and Boosuyt, 2003
SIA	3	1st & 2nd exon	3rd exon	Bonacum et al, 2000
Rag1	2	1st exon	2nd exon	Biju and Boosuyt, 2003
Rag2	3	3rd exon	1st exon	Hoegg et al, 2004
POMC	3	1st exon	3rd exon	Wiens et al, 2005
SLC8a3	6	---	1st exon	Shimada et al, 2011
NCX1	10	---	1st exon	Shimada et al, 2011
BDNF	2	---	2nd exon	Meijden et al, 2007

Nuclear gene fragment

Our objective was to sequence the region between (and including) two exons – in a gene used in previous studies for phylogeny (Table 2.2) – which were separated by 5-10 kb. We expected the intronic regions, which are likely to be under far less selection pressure than protein-coding exons, to show significantly higher mutation rates and hence a stronger signal to study genetic structure at the intra- and inter- population levels.

Table 2.2 shows the candidate genes that we considered for this exercise. It lists the number of exons, untranslated regions (UTRs – to avoid locating primers in them) if any, and the primer locations. We selected Rhodopsin as the target gene since (i) NCBI sequences were available for both the 1st and the 4th exons (X1 and X4) for locating the forward and reverse primers, and (ii) the DNA length between them provided about 6000 intronic bases – perhaps long enough for separating close taxa but not so long as to cause issues with long-range PCR. We also considered the ~10 kb intergenic region between Rag2 (3rd exon) and Rag1 (1st exon) but decided to go with the shorter Rhodopsin to avoid the more stringent quality requirement of PCR amplification of the longer fragment.

While there were many anuran sequences of the 1st and 4th exon (X1 and X4), they could not provide the length of the region between them. This could only be obtained from species for which the entire gene had been assembled and annotated. When this project was first envisaged the whole genome sequence was available for only 2 anuran species: *Xenopus tropicalis* (Hellsten et al. 2010) and *Nanorana parkeri* (Sun et al. 2015) The annotated gene of *X. tropicalis* yielded a length of 2318 bp for Rhodopsin X1-X4. We identified X1 and X4 on (a contig of) the *Nanorana* genome and obtained a length of 5494 bp.

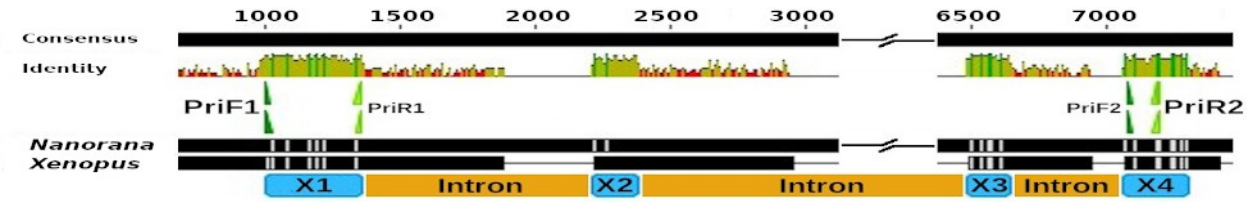


Figure 2.6: Alignment of *Xenopus* and *Nanorana* sequences. Genome contigs are thick black bars, exons are blue, introns are orange, primers are dark green (forward) & light green (reverse). Histogram on top indicates extent of match between sequences (green: match, red: no match). Exon order is in forward sense 1st (X1: left-most) to 4th (X4: right-most).

We designed two pairs of forward and reverse primers for both X1 and X4: FRH-1A and FRH-1B, forward primers for X1; FRH-1C and FRH-1D reverse primers for X1; FRH-4A and FRH-4B, forward primers for X4; and FRH-4C and FRH-4D reverse primers for X4. These primers are partially modified versions of the ones published by Bossuyt and Milinkovitch (2000). A PCR reaction of the within-exon primers (e.g. FRH-1A and FRH-1C in X1; FRH-4A and FRH-4D in X4) was conducted to first confirm the binding and specificity of all primers with short fragment under identical conditions. Subsequently the forward primer of X1 (FRH-1A) and the reverse primer of X4 (FRH-4D) were used to amplify the entire fragment of ~6000 bp in one long-range PCR.

The primers were initially tested on individuals from 3 different frog families. Figure 2.12 shows successful amplification using the primers of species from the families Dicroglossidae, Rhacophoridae and Ranidae respectively.

Section 2.2.4: PCR amplification

We used PrimeStar GXL (Catalog R050A) polymerase from TAKARA for the long-range PCR. PrimeStar GXL is a high fidelity PCR enzyme which can generate amplicons of 30 kb or higher. We also added Bovine Serum Albumin (BSA) at 5 mg/mL to the mix to prevent reagents from sticking to the walls of the vials. We carried out the 20 µL PCR reactions in 0.2 mL vials. The PCR master mix was prepared by adding the required volumes of all the reagents, with the polymerase (stored on ice on the bench) being the last to be added.

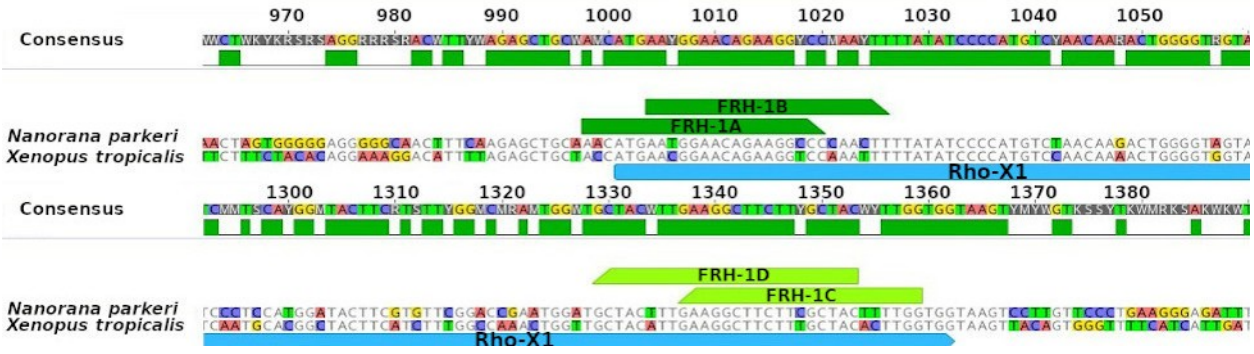


Figure 2.7: Location of primers on exon 1 (X1) of the Rhodopsin gene. FRH-1A and 1B are the forward primers and FRH-1C and 1D are the reverse primers.

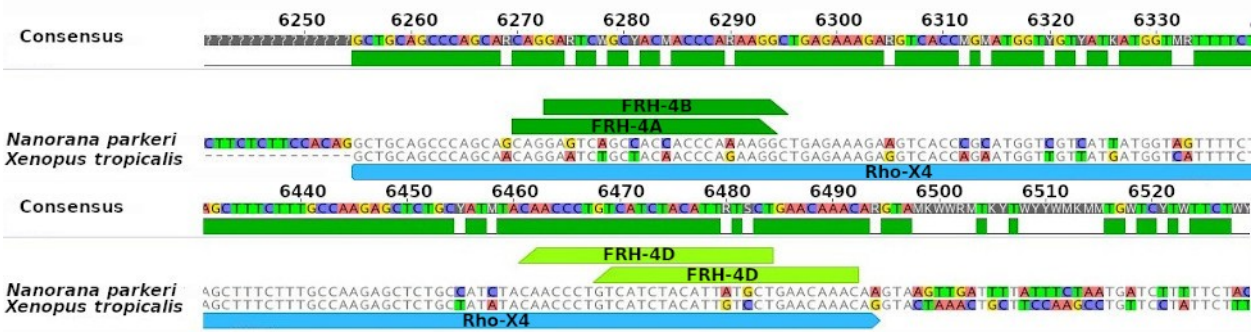


Figure 2.8: Location of primers on exon 4 (X4) of the Rhodopsin gene. FRH-4A and 4B are the forward primers and FRH-4C and 4D are the reverse primers.

The reaction mix is listed in Table 2.3. The reaction parameters were (i) Denaturation: 98 °C, 10 s (ii) Annealing 55 °C, 15 s, and (iii) Extension: 68 °C, 1 min per kb. The list of primers designed and/or used in this study is listed in Table 2.5.

Extension time: After testing several pairs of mitochondrial primers, we used the pair FMT1-TF and FMC2-TR since they yielded the longest product of 7178 bp. The two reference genomes yielded nuclear Rhodopsin gene (X1 to X4) lengths of 5493 bp (*Nanorana*) and 2318 bp length in *Xenopus*. Intron length can vary a lot between species, and too short a time results in no amplification at all. Therefore, to be on the safe side we set 10 min extension time (sufficient for up to 10 kb length) for all reactions.

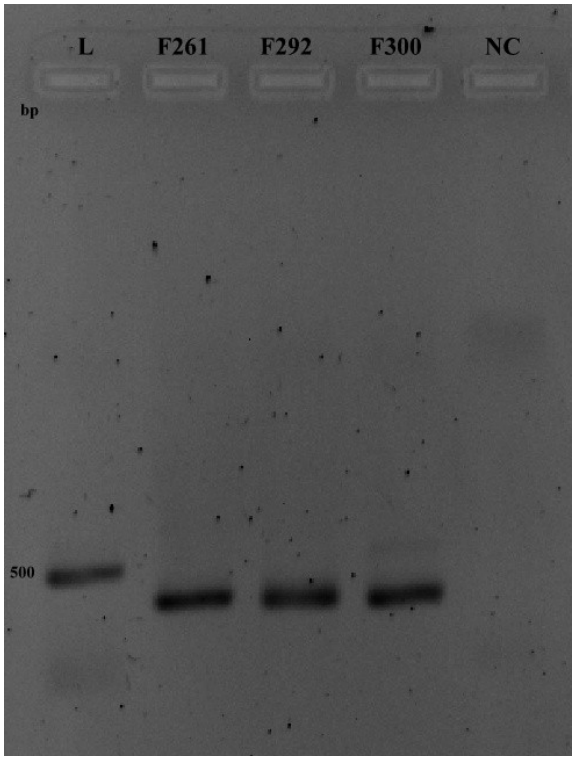


Figure 2.9: Electrophoresis image demonstrating the binding of the primers using short-range, within-exon PCR. The labels are L: ladder (with a single 500 bp marker), NC: negative control and 3 frog samples (in order, *Dicroglossidae, Rhacophoridae*, and *Ranidae*

We confirmed that the long-range PCR of Rhodopsin (1st to 4th exon) worked for multiple genera including *Occidozyga*, *Xenophrys*, *Amolops*, *Nanorana*, *Chiromantis*, *Kurixalus*, *Raorchestes*, *Fejervarya*, *Amolops*, *Theloderma*, *Leptobrachium* and *Polypedates*.

In this study, we will only discuss the analysis related to *Raorchestes* mitochondrial samples.

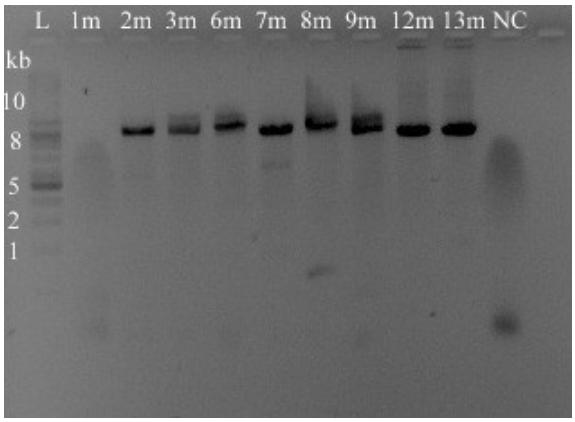


Figure 2.10: Electrophoresis of *Raorchestes* mitochondrial fragment for different extension times.

Section 2.2.5: Index PCR

Multiplexing hundreds of samples was made possible using barcode indices at three different locations. Nanopore library prep involves adding a barcode index. The sequencer software automatically identifies this index in the reads and sorts the output according to this index.

Additionally, we used 10 internal primers in the regions before the forward amplicon primer (IndexF in Figure 2.11), and 10 internal primers in the region after the reverse amplicon primer (IndexR in Figure 2.11). These corresponds to 10 x 10 = 100 unique combination. We added these barcode indices using a second PCR which were primed using the T7 and T3 tails in the primers of the amplicon (first) PCR. The additional bases K-F and K-R shown in Figure 2.11 are not of relevance to this exercise. They had been used in a previous study to add sequencing flow cell adaptors to the fragments.

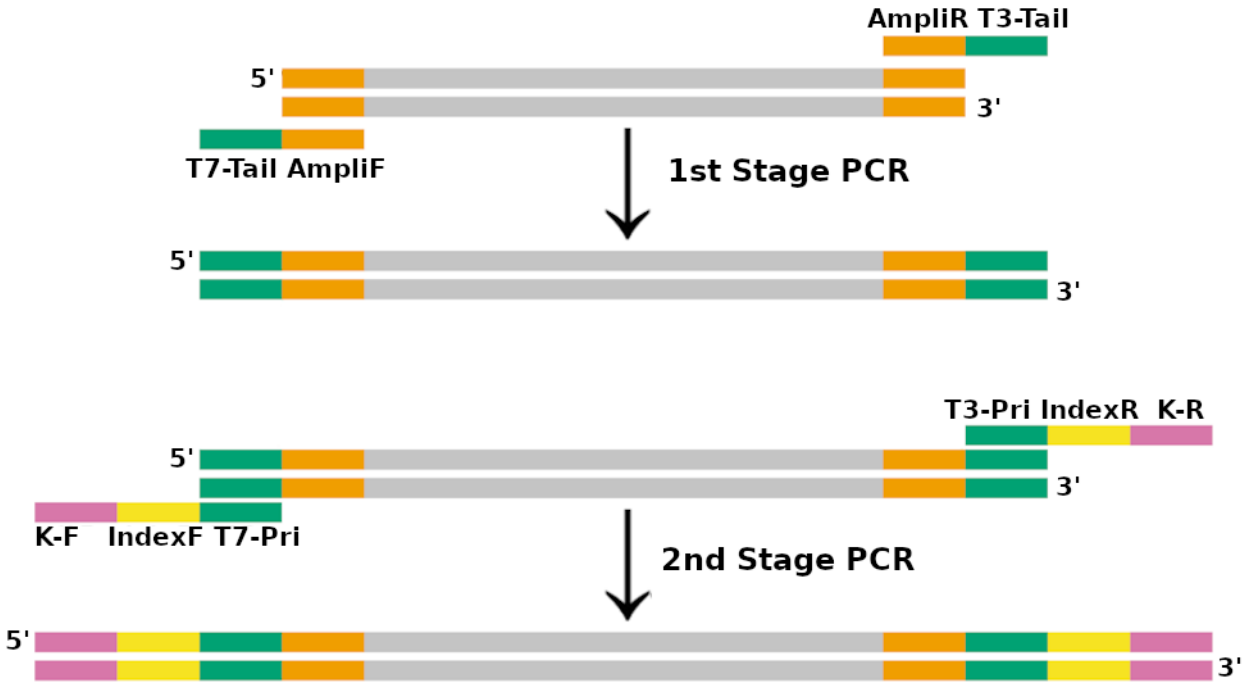


Figure 2.11: 2-stage PCR for amplicon generation and indexing. AmpliF and AmpliR are the amplicon primers. T7 and T3 promoter sequences were used as the tail in the first PCR and primers for the second PCR to add the barcode indices IndexF and indexR. K-R and K-F are not of significance for this study, being carry-overs from a previous study which had used the same barcode indices.

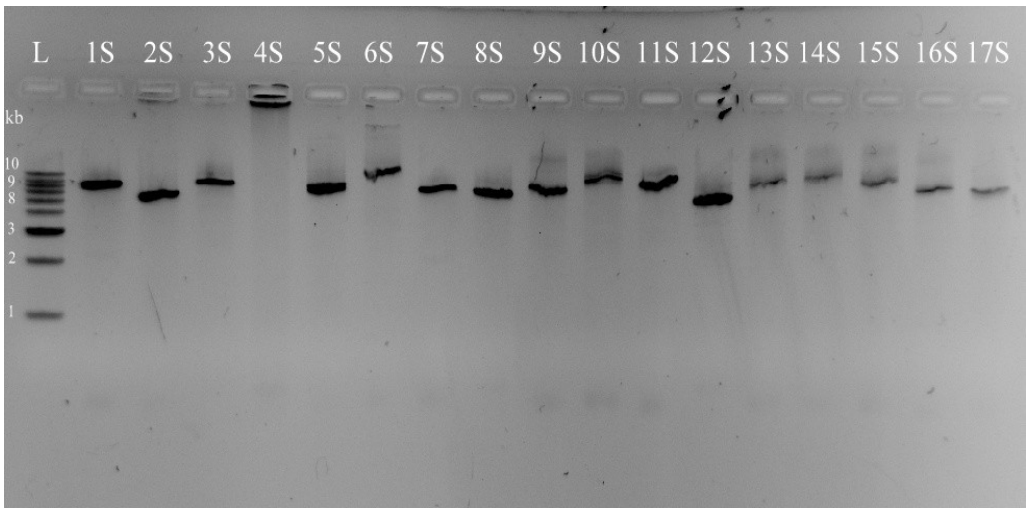


Figure 2.12: Electrophoresis image of the PCR products of the nuclear rhodopsin gene for multiple genera including *Occidozyga*, *Xenophrys*, *Amolops*, *Nanorana*, *Chiromantis*, *Kurixalus*, *Raorchestes*, *Fejervarya* (11S), *Amolops*, *Theloderma*, *Leptobrachium* and *Polypedates*. The bands varied in size but were all in excess of 6 kb.

Table 2.3: Summary of all PCR reagents, their initial and final concentrations

Reagents	Reagent Initial conc	Volume per reaction (µL)	Reagent Final conc
PrimeStar GXL polymerase	1.25 (U/µL)	0.4	0.025 (U/µL)
Buffer	5X	4	1X
dNTP mix	2.5 mM each	1.6	0.2 mM each
Primer F	10 µM	0.5	0.25 µM
Primer R	10 µM	0.5	0.25 µM
BSA	5 mg/mL	3.4	0.85 µg/µL
Water	---	7.6	---
Template DNA	10-100 ng/µL	2	1-10 ng/µL

Table 2.4: Amplicon primers used in this study. Primer tails (T7 and T3) are listed at the end

Primer Name	Locus	Sequence	Modified from
FMT1-TF	mtDNA-12S	T7-CTTAAAACYCAAAGGAYTTGWCGGT	Darst & Cannatella. 2003
FMC2-TR	mtDNA-tRNALys	T3-CACCTRYCTACAGCTTAAAAGG	This study
FRH-1A	Rho-Exon1	T7-ACCATGAAYGGAACAGAAGGYCC	Bossuyt et al. 2000
FRH-1B	Rho-Exon1	T7-AACGGAACAGAAGGYCCMAAYTT	Bossuyt et al. 2000
FRH-4C	Rho-Exon4	T3-TTTGTTCAGSAYAATGTAGATGAC	Bossuyt et al. 2000
FRH-4D	Rho-Exon4	T3-AGSAYAATGTAGATGACRGGGTTG	Bossuyt et al. 2000
T7: TAATACGACTCACTATAGGG		T3: CAATTAACCCTCACTAAAG	

Table 2.5: Barcode indices with tails (K1, K2, K3) and PCR priming sections (T3 and T7). The tail sequences were incidental (being designed for another project) and not of consequence.

Name	Forward Index Sequence		Name	Reverse Index Sequence
T7F1	**K1-ACGCTGTACGCG-T7**		T3-F1	**K3-ACTCGTGACT-T3**
T7F2	**K1-CATAGTGCATACG-T7**		T3-F2	**K3-CAGATGTCAGA-T3**
T7F3	**K2-GTATCACG-T7**		T3-F3	**K2-GTCTACAGTCTG-T3**
T7F4	**K1-TGCGACATGCG-T7**		T3-F4	**K3-TGAGCACT-T3**
T7F5	**K2-AGTGCTCA-T7**		T3-F5	**K2-AGTGCTCA-T3**
T7F6	**K2-CTGTAGAC-T7**		T3-F6	**K2-CTGTAGAC-T3**
T7F7	**K2-GACATCTG-T7**		T3-F7	**K2-GACATCTG-T3**
T7F8	**K2-TCACGAGT-T7**		T3-F8	**K2-TCACGAGT-T3**
T7F9	**K2-ATCTGCGA-T7**		T3-F9	**K2-ATCTGCGA-T3**
T7F10	**K2-CGAGTATC-T7**		T3-F10	**K2-CGAGTATC-T3**
K1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT		**K2**	CGATCT
K3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT			

Section 2.2.6: PCR Product Purification

Several techniques are available for purification of PCR products including silica column based filters with molecular size cut-off and bead based purification. Our primary considerations, apart from elimination of unused primers and other contaminants, was intact large fragment of DNA for downstream processing and expense. We selected the MAGBIO HighPrep PCR Clean-up kit (AC-60005) which uses magnetic bead based purification and is less expensive than other PCR clean-up systems. This bead-based procedure selects fragments of a particular size based on the ratio of the volume of bead mix and PCR reaction, with lower bead ratio selecting longer DNA fragments. Following the manufacturer's protocol we tested the purification using bead volumes of 4 µL, 8 µL, and 18 µL with 10 µL PCR product. The reaction with 0.4X bead volume resulted in a good degree of elimination of unused primers and primer-dimers (Figure 2.13). Figure 2.14 provides a comparison of the pre- and post-cleanup PCR products – the post-cleanup lanes are free of primer-dimers and unused primer signals.

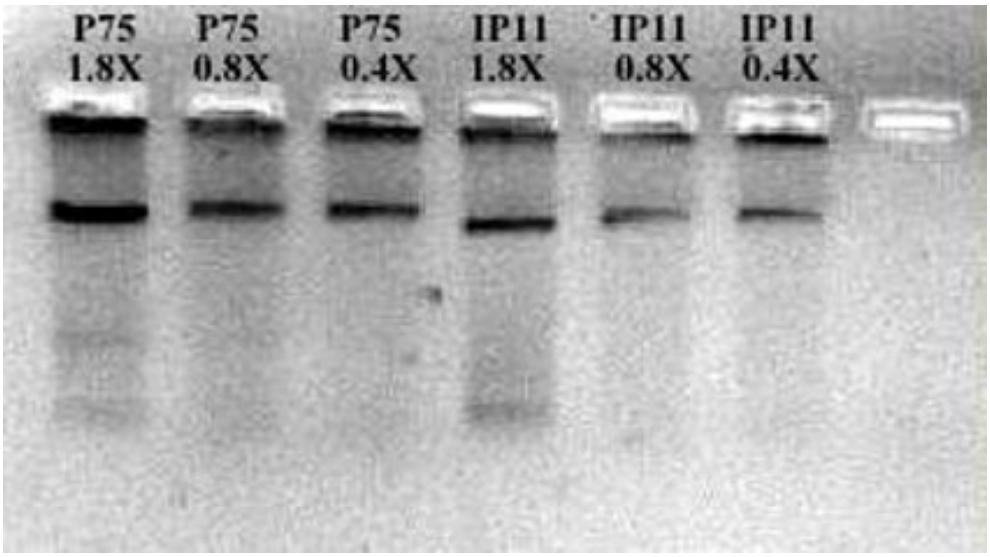
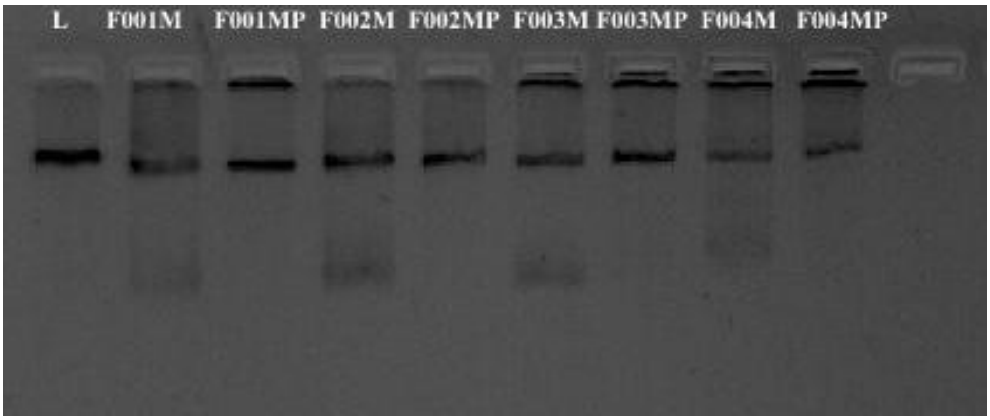


Figure 2.13: Electrophoresis image of the results of bead purification protocol. Sets of PCR products were purified with bead volume ratios of 1.8, 0.8 and 0.4.



F**igure 2.14**: Comparison of pre- and post-purification PCR product with 0.4X bead volume. The samples with the suffix M and MP are pre- and post-purification, respectively.

Section 2.2.7: DNA Quantification

Measuring DNA concentration with reasonable accuracy is important for creating an equimolar pool of multiple samples for Next Generation Sequencing (NGS). One can measure DNA either using absorption based spectrophotometer (e.g. NanoDrop) or with an emission based fluorometer.

An absorption spectrophotometer, e.g. NanoDrop, is useful to obtain the composite absorption spectrum of all the constituents of a sample. The absorption at the frequency associated with the DNA may have contributions from other constituents in the sample. Furthermore, though a

NanoDrop does not require any additional reagents (and hence is very inexpensive), it is very labour intensive and unsuited for processing a large number of samples.

DNA does not fluoresce by itself, but fluorescence may be observed by a bound complex of DNA and some organic dyes (e.g. Ethidium bromide, SYBR etc). This technique can be very sensitive and accurate and may even be used to detect a single molecule (Valeur B and Berberan-Santos M.N. 2013, Simbolo et al. 2013). We rejected the use of Qubit fluorometer (Haendiges et al. 2020) to quantify DNA as it is both expensive and labour intensive. DNA can also be detected using PCR in the presence of fluorophore-linked single-stranded oligonucleotides with high specificity of binding to the amplicon of interest. This technique lends itself to high accuracy and high throughput for a large number of samples but is very expensive.

Instead, we used the TECAN Infinite M200 Pro, a micro-plate reader to measure the concentration of DNA in our purified PCR products. This procedure provides quite accurate measurements (error: ~1 ng/uL) but comes with the advantage of processing upwards of 64 samples at a time and is extremely inexpensive (approx. 1 USD for a plate of 64 samples).

We note that pooling multiple DNA samples for NGS only requires knowledge of the relative DNA concentration at the level of the individual samples. We created a set of standard DNA samples of 10, 20, 30, 40, 50, 75 and 100 ng/μL by dilution from a readily available high DNA concentration source like salmon sperm (Sambrook and Russell. 2001) or mouse tissue extracted in the lab. The

Table 2.6: Configuration of well concentrations in a 96-well plate reader used to measure linearity of Tecan Infinite M200 Pro. The values shown are the concentration of the DNA standard set.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
1	100	100	100	0	100	100	100	0	100	100	100	100
2	75	75	75	0	75	75	75	0	75	75	75	75
3	50	50	50	0	50	50	50	0	50	50	50	50
4	40	40	40	0	40	40	40	0	40	40	40	40
5	30	30	30	0	30	30	30	0	30	30	30	30
6	20	20	20	0	20	20	20	0	20	20	20	20
7	10	10	10	0	10	10	10	0	10	10	10	10
8	0	0	0	0	0	0	0	0	0	0	0	0

concentration of this purified standard set was measured using the NanoDrop. The concentration of all samples were determined in relation to the response of the fluorometer for this reference set.

We note that pooling multiple DNA samples for NGS only requires knowledge of the relative DNA concentration at the level of the individual samples. We created a set of standard DNA samples of 10, 20, 30, 40, 50, 75 and 100 ng/µL by dilution from a readily available high DNA concentration source like salmon sperm (Sambrook and Russell. 2001) or mouse tissue extracted in the lab. The concentration of this purified standard set was measured using the NanoDrop. The concentration of all samples were determined in relation to the response of the fluorometer for this reference set.

As a first step we measured the linearity of the response of TECAN by measuring the fluorescence counts for the standard set. Each well was loaded with a mix of 99 uL 1X SYBR Gold and 1 uL of DNA in 1X-TE. The configuration of the plate reader for this assay is shown in Table 2.6. The result of the experiment is plotted in Figure 2.15. The linearity is quite good to better than 10% out to 100 ng/uL. Though, an input value of ~ 20-30 ng/uL seems to be best in terms of dispersion.

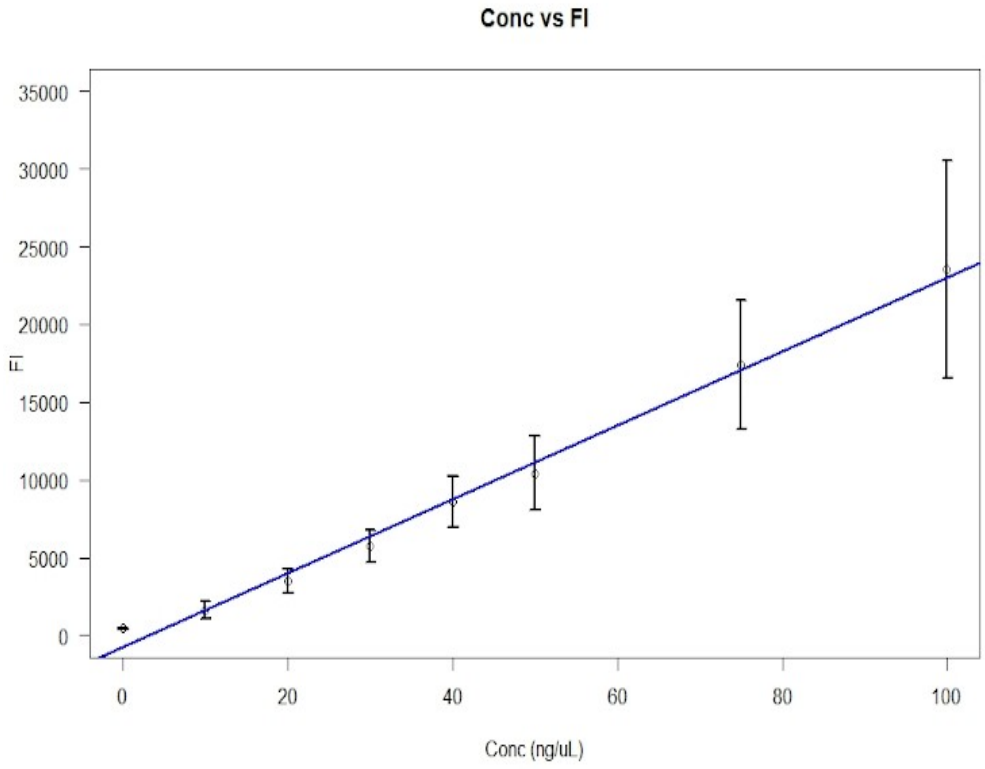


Figure 2.15: Linearity of TECAN Infinite M200 PRO. X-axis is the input concentration of the standard set and Y-axis is the fluorescence counts.

There were two further issues to be solved for using this method:

1. non-zero fluorescence for zero DNA input: which would result in considerable overestimation of the DNA concentration
2. Measurement of the unknown sample concentration with optimal measurement concentration of 20-30 ng/uL: clearly a problem which require an iterative solution

Both these issues were solved using an advantage of a plate-reader that an instrument like the Qubit does not have. One issue with Qubit is that the sample and the standard have to be measured under approximately same conditions of temperature and exposure to ambient light, since the DNA binding fluorescent dye (like SYBR) usually degrades when exposed to light. This can be an issue when dealing with a large number of samples.

The default for a 96-well plate is that all the samples are exposed to the same ambient light and temperature. Furthermore, TECAN only take about 3 minutes to read all the wells while the plate is held in a controlled environment. Also, the same plate can be inserted into the instrument and read multiple times – while the fluorescence counts may change from one insertion to the next, all the samples experience the same ambient condition and hence their relative fluorescent response does not change.

Non-zero fluorescence of the well: We observed that the background (i.e. counts without DNA) fluorescent intensity varied from well to well, from plate to plate, and with time. However, the relative values across the wells remained the same for a plate, even across several weeks. Therefore, we followed the following protocol:

1. The schematic for measuring the DNA concentration is shown in Table 2.7
2. Initially fill all the wells with 99 µL of 1X-SYBR and 1 µL of 1X-TE (total 100 µL)
3. Measure the fluorescence of the wells
4. Take out the plate and add a further 1 µL to each well, consisting of
    i. 1X-TE without any DNA in the wells Z1
    ii. 20 ng/µL DNA standard into C20
    iii. 30 ng/µL DNA standard into C30
    iv. 40 ng/µL DNA into C40
    v. target DNA to be measured into the wells labelled DNA-binding

5. Measure the plate again.

The mean ratio of Z0 and Z1 provided a measurement of gain variation between measurements 1 and 2. As previously mentioned, the relative variation across the wells remained constant over weeks. We then did a well-by-well subtraction of the background using the different values of Z0 and the mean ratio of Z1/Z0. The gain function of the instrument – in counts per nanogram – was calculated using the wells with standard DNA (C20, C30 and C40). This was used to determine the concentration of the target DNA.

We note that

1. the relative numbers of the different categories of wells may be changed as per convenience and requirements of efficiency and accuracy.
2. We loaded each target sample into 2 wells to confirm accuracy of the values. This may also be increased or decreased as per requirement.
3. We set the instrument to measure each well 5 times. This may also be changed as per requirement.

Table 2.7: Plate layout for measuring DNA using a plate-reader. Z0 are measurements without DNA (99 uL 1X-SYBR + 1 uL 1X-TE). Subsequently, 1 uL was added to the wells for Measurement 2 of the same plate: 1X-TE to Z1, standard 20 ng/uL DNA to C20, standard 30 ng/uL DNA to C30 and standard 40 ng/uL DNA to C40 and DNA samples in the rest.

Measurement 1

	1	2	3	4	5	6	7	8
A	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0
B	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0
C	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0
D	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0
E	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0
F	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0
G	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0
H	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0
I	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0
J	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0
K	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0
L	Z0	Z0	Z0	Z0	Z0	Z0	Z0	Z0

Measurement 2

1	2	3	4	5	6	7	8
C20	DNA	DNA	C40	DNA	DNA	DNA	DNA
DNA	DNA	DNA	DNA	DNA	DNA	C20	DNA
DNA	DNA	C20	Z1	C40	DNA	DNA	DNA
DNA	C30	DNA	Z1	DNA	C20	DNA	DNA
C40	DNA	DNA	Z1	DNA	DNA	DNA	C30
Z1	Z1	Z1	Z1	Z1	Z1	Z1	DNA
DNA	DNA	DNA	Z1	DNA	DNA	DNA	DNA
DNA	DNA	C30	Z1	DNA	DNA	C30	DNA
DNA	C20	DNA	Z1	DNA	DNA	DNA	DNA
DNA	DNA	C40	Z1	DNA	C20	DNA	DNA
C30	DNA	DNA	DNA	C30	DNA	DNA	DNA
DNA	C40	DNA	DNA	DNA	DNA	DNA	C40

Achieving optimal concentration for measurement: If a measurement indicated a higher concentration than desired for a target, we simply diluted the well with SYBR+TE solution by the appropriate amount and re-measured the plate. On the other hand if the measured concentration was lower than optimal we simply added more of the target DNA into the well and compensated for this during the calculation. We observed that the addition of 10 μL into a well, and the consequent change of the geometry of the liquid in the well did not appreciably change the instrumental counts.

Section 2.2.8: Next Generation Sequencing (NGS)

We explored two NGS platforms to sequence the 350 samples of length 6000-7000 bp. The first was Illumina using tagmentation to prepare the library and the second was Oxford Nanopore.

Library-prep for Illumina using Tagmentation

We used Illumina MiSeq as the platform for the initial attempt at sequencing the DNA fragments. It produces paired end reads of 150 bp each. Illumina requires fragmentation of the target region into reads of size 200-500 bp. The target sequence is determined by reassembling the reads. Illumina reads have to be labelled with identification barcodes to be able to separate those from different samples. The technique of Tagmentation, fragments the target amplicon and simultaneously tags them with adaptors. A PCR may then be performed on these fragments with the adaptors serving as primer binding regions to add identification indexes (barcodes) and external Illumina sequencing flowcell adaptors to all the fragments of a sample. As it happens, while the actual cost of sequencing the reads is low, the cost of sample preparation is 10-100 times higher. This makes it prohibitive for a study which requires pooling PCR amplicons from several hundred samples. We experimented with two strategies to reduce the cost:

1. Development of in-house protocol which would be a substitute for commercial tagmentation
2. Standardise use of commercial tagmentation reagents at much lower volume than recommended

Illumina's commercially available Nextera kits makes use of transposition for tagmentation but the structure of their protein as well as other reagents are undisclosed.

Picelli et al (2014) described procedures for cost-effective tagmentation based library preparation using in-house Tn5-transposase. Transposase is a protein that facilitates excision of DNA from donor sequence and inserts it into the target sequence. The paper further demonstrated how to modulate library lengths and generate tagmented library from subpicogram levels of DNA. We synthesised the enzyme in the lab and succeeded in tagmenting the target DNA. However, we abandoned this exercise since it was felt that getting stable performance from the in-house enzyme required more time than at our disposal.

We also explored an alternative strategy of reducing the volume of DNA, and hence tagmentase per sample to reduce the cost. We achieved routine success while using just 1/3rd of the enzyme recommended in the protocol, but this was still very expensive from our perspective. We targeted a reduction of the volume to 1/10th and even 1/20th. We achieved some degree of success and in fact sequenced 10 samples. However, consistency of tagmentation quality was still an issue at those low enzyme levels. More importantly, we found that tagmentation resulted in considerable depth bias across the target fragment. Further, reconstructing the sequence especially in the intronic region (with repeated sections) of non-standard organisms was an issue.

Therefore, we settled on the Oxford Nanopore sequencing platform which provided an easy and inexpensive method of library prep, and also long fragment sequencing which did not have issues of assembly.

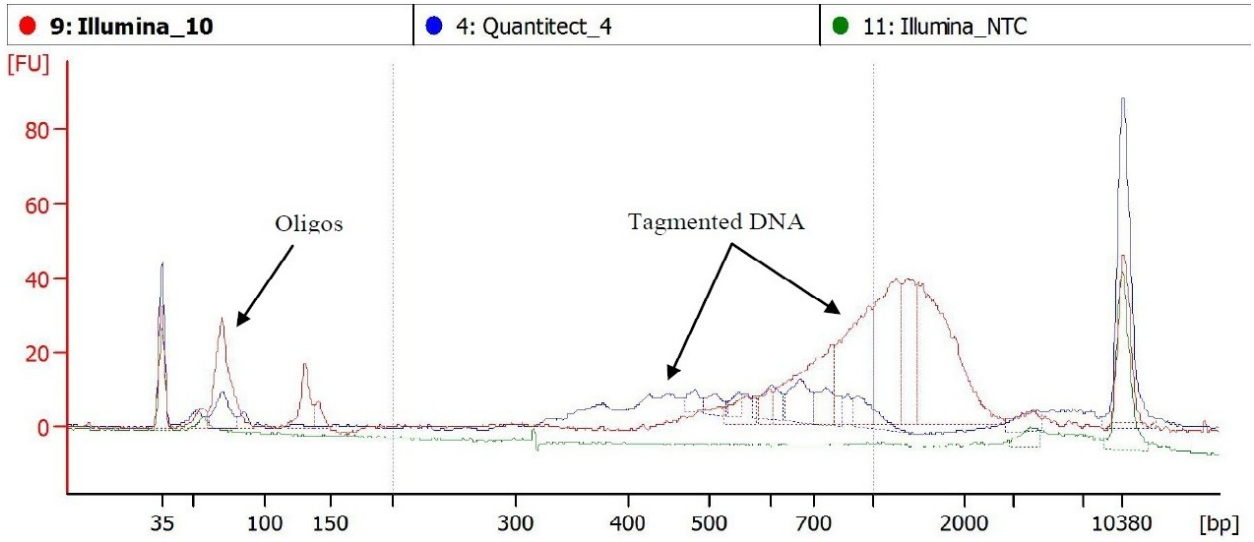


Figure 2.16: Bioanalyzer plot of the DNA size (bp) profile after tagmentation by the in-house synthesised tagmentase. The tagmented DNA was subsequently barcoded via PCR using both Quantitect and Illumina's polymerase. The Quantitect product in particular shows considerable yield. NTC is the plot of the DNA-less negative control.

Library-prep for Nanopore MinION

The availability of a third generation sequencing platform, MinION, from Oxford Nanopore Technologies (ONT) persuaded us to alter our library preparation techniques for sequencing on MinION. The MinION device is a small portable device of dimensions 10 x 3 x 2 cm and weighing only 90 gm. It can be plugged directly into a standard USB3 port on a computer with modest hardware requirements, as low as 8 GB RAM and a solid-state storage disk of at least 128 GB (recommended 1 Tb). The MinKnow software which allows users to control various aspects of sequencing can run on any operating system.

The principal steps during library prep include the addition of identifying indices (barcodes) for each sample, followed by sequencing flow cell adaptors. In the case of Illumina, which can only work with short fragments, this required (expensive) tagmentation of each of the hundreds of samples, which increased the cost a lot.

MinION works best with long DNA sequences and so fragmentation was not needed. We added the barcodes using a PCR, rather than the more expensive tagmentation. Subsequently, the flow cell adaptor for the entire pool was added using a single ligation reaction.We split the set of samples into

3 pools. Each individual in a pool had a unique barcode combination. The first pool comprised 100 indexed samples of *Raorchestes* mtDNA and nuDNA and 50 of *Megophrys*, the second pool comprised the next 100 samples of each category, and the third pool contained the remaining samples. Since our samples varied considerably in concentration, in each pool, we sorted the samples into subclasses of DNA concentration: [1.0, 2.0], (2.0-5.0], (5.0-10.0], (10.0-20.0], (20-50], and (50-150]. We pooled the DNA sample within each subclass in equimolar proportions. The concentration of the sub-pools were measured using Tecan and the calculated concentrations were found to be similar to expected ones. Finally, the sub-pools were mixed in such proportions that all the samples in a pool were equimolar. This hierarchical mixing was undertaken to avoid dealing with small volumes.

To summarise, at this stage we had 3 pools of equimolar DNA samples consisting of mtDNA and nuDNA fragments. The first two pools had 100 unique barcode combinations representing 100 unique individuals. We note that these 100 barcode combinations were the ones we had attached

Table 2.8: Final DNA volume to be added per pool

Label	Number of samples	Pool conc (ng/µL)	Volume (µL)	DNA (ng)
Pool1	251	5.26	25	131.5
Pool2	258	4.66	25	116.5
Pool3	106	5.92	16.8	99.5

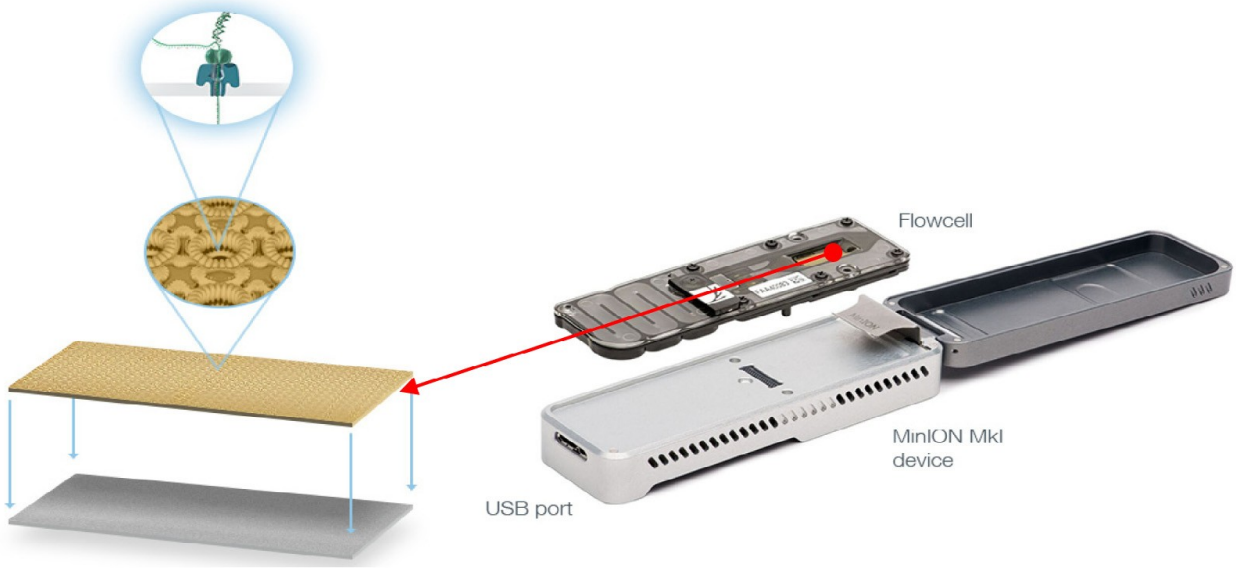


Figure 2.17: MinION device showing the flowcell and membrane pores embedded inside it (Lu et al, 2016)

using the PCR, and were the same in all the pools. In the next step we ligated the Nanopore barcodes and flowcell adaptors so that the 3 pools could be sequenced in the same run.

We followed the manufacturer's protocol specifically designed for sequencing amplicons. Nanopore barcodes were attached to the three pools using 3 (of the 12) barcodes in the kit EXP-NBD104. The other reagents were from SQK-LSK109 kit. The procedure included:

- **Library end-prep and dA tailing:** Each pool (28 µL) was mixed with nuclease free water (20 µL), NEB Ultra II End-prep reaction buffer (3.5 µL), NEB Ultra II End-prep enzyme mix (3.0 µL) and incubated at 20°C for 5 min followed by 65°C for 5 min. This step repaired any sticky ends (unlikely for amplicons) and added the nucleotide "A" to the ends of each fragment. This end-repaired library was purified using Ampure XP beads.
- **Barcode ligation:** Nanopore barcodes NB01, NB02 and NB03 were ligated to the 3 pools by mixing the barcodes (2.5 µL for each) with end-repaired library (22.5 µL) and NEB Blunt TA ligase (25 µL) and incubating for 10 min at room temperature. The products were again purified with Ampure XP beads.
- **Equimolar pooling of barcode attached libraries:** The details of each pool are listed in Table 2.8. The Nanopore barcoded pools were mixed to obtain equimolar concentrations of all fragments (proportional to the number of amplicons in each) … flow cell pool.
- **Sequencing adapter ligation:** 65 µL of the flow cell pool was added to Adapter Mix II (5 µL), NEBNext quick ligation buffer (20 µL), and quick T4 DNA ligase (10 µL) and incubated at room temperature for 10 min. Post-adapter ligation, the pool was purified with Ampure XP beads and eluted in 15 µL nuclease free water.

The entire process yielded an NGS library of concentration 15.1 ng/µL, which is ~48.70 fmol (amplicon size ~7 kb). The recommended concentration for loading on the flow cell is 5-50 fmol. We loaded on the flow cell a mix of 7 µL library, 5 µL nuclease free water (effectively 12 µL library of 20.3 fmol), sequencing buffer (37.5 µL) and loading beads 25.5 µL.

Sequence Data

The sequencer was run for 48 hrs to obtain ~10 Gb (giga bases) of data. The manufacturer's software MinKnow outputs both raw fast5 and base-called fastq files. This software also separated the data for the 3 Nanopore barcodes: NB01, NB02 and NB03. We obtained 833, 1220 and 492

fastq files comprising 7.53 GB, 8.15 GB and 4.62 GB respectively (note: fastq contains both bases and quality, hence size would be double the number of bases sequenced; GB=Gigabases). Each fastq file contained 4000 reads except the last one of each barcode. Overall the run yielded 3331007, 4879671 and 1967352 reads, respectively.

Demultiplexing Reads: separating the internal barcodes

Read sequences of individual samples were demultiplexed first based on their amplicon sequences followed by the index sequence. All the reads of each (internal) barcode along with the read-names were extracted to one csv file.

Subsequent analyses were performed individually for each sample, using scripts written in R (v3.6.2).

We used clustalW from the package *msa* to search for the amplicon primer sequence (without the tail) in the first 200 bases of a read. The total length of flowcell adaptor + index + amplicon was only ~66 bases. Therefore, restricting the search to 200 bases was deemed sufficient and considerably reduced the time needed for computation. Since the 5' end of the read could either have the forward or reverse amplicon primers, we searched for both. A match flag was set if the number of mismatches was 0-3 bases. The forward and reverse amplicon matches were stored in separate files.The matched reads of each primer pair were next searched for the identity of index sequences: 10 indices with T7 tail with the forward amplicon primer and 10 indices with T3 tail with the reverse amplicon primer, a total of 100 combinations. The matching of the barcode index was tested only in the immediate region before the amplicon+T7/T3 locus at the 5' end or in the immediate region after the amplicon+T7/T3 locus at the 3' end. Barcode index flags were set to match only if the number of mismatches was 2 or fewer. Only a read with matching indices on both the 3' and 5' ends were considered as valid, and assigned to the appropriate individual frog.

All the reads of an individual were aligned in Geneious (v2022.2.2). Even at this stage the aligned reads were inspected by eye to remove obvious alignment errors. Any erroneously assigned read which either had wrong index or more than two mismatches were removed.

The calling of the consensus base at a locus is the subject matter of the next chapter.

Chapter 3

A Consensus Base-Calling Procedure for NGS Data

Section 3.1: Introduction

Advancement in sequencing strategies both in terms of computational and technological improvement have found numerous applications in ecology such as biodiversity assessment through metabarcoding (Taberlet et al., 2012; Deagle et al., 2019), profiling microbiome communities (Fierer, 2017), understanding population structure for conservation implications (Allendorf et al., 2010), phylogeographic analysis (Robin et al., 2010) to name a few. The limited multiplexing capabilities of Sanger sequencing (Sanger and Coulson, 1975) was greatly improved by the introduction of next generation sequencing technologies which enabled high-throughput generation of millions of reads while simultaneously sequencing hundreds of samples (Shokralla et al., 2014; Golan et al., 2012).

Next generation sequencing technologies like Illumina and Roche rely on DNA fragmentation followed by assembly of the sequenced fragments (about 150-300 bases) to obtain the sequence of entire region of interest. Illumina, in particular, provides reads which are inexpensive (cost per base) and accurate. However, as mentioned in the previous chapter, the library-prep is expensive for highly multiplexed targets and the assembly can be problematic while resolving repeat regions, detecting structural variant (Stancu et al., 2017). Biases in fragmentation can also result in uneven coverage of the region of interest.

The development of single molecule real time (SMRT) sequencing implemented in Pacbio and Nanopore technologies address the problem of fragmentation and assembly. These technologies allow sequencing of an entire intact single DNA molecule of over 800 kb (Jain et al., 2018) even exceeding 2 Mb (Payne et al., 2019) in real time. The MinION platform implementing nanopore technology and commercialised by Oxford Nanopore Technology (ONT) is a portable sequencing platform which does not require a traditional dedicated sequencing facility – any computer with modest hardware configuration can become a sequencer with the MinION.

The major limitation of Nanopore sequencing, compared to Illumina, is the high error rates. MinION, when it was introduced, provided a low read accuracy of 60% (Goodwin et al., 2015, Laver et al., 2015) which gradually improved to 85% (Jain et al., 2018; Stancu et al., 2017; Jain et al., 2017; Tyson et al., 2018) but still falls short of the >99% accuracy offered by Illumina. Read accuracy has been commonly defined as the percentage of bases in a read segment that match the reference out of the total length of the read segment-reference alignment (Rang et al, 2018). The reported per read accuracy is also based on the subset of reads that can be aligned to a reference sequence (Rang et al., 2018). Although the average accuracy per read is low, the consensus generated from homologous read alignment has reported final accuracies exceeding 99% (Loman et al., 2015; Li H. 2016; Koren et al., 2017; Wick et al., 2018). While a final accuracy of 99% (1 in 100 error) can be sufficient in many cases, we set out to achieve accuracies of <1 in 1000 to be able to construct well-resolved phylogenetic trees at the intraspecific level and to detect population diversity on time scales much smaller than climate cycle periods.

When sequencer reads are aligned to a reference one usually see bases of more than one kind at each locus. One needs an objective criterion for calling a particular base as the consensus at a locus. Sequencers usually assign a quality to each base in a read in terms of the Phred score

$$Q = -10 \log (\text{probability of a base call being incorrect})$$

The Q score has often been used as a reliability metric in identifying the consensus base at a locus.

We present here a list of issues that we have identified from published literature (e.g. (Bonfield & Staden 1995; Azam et al. 2012; Stock et al. 2024; Espada et al. 2022):

1. Given the centrality of reliability of a base (i.e. its Q score) to consensus-call, how close is the formal error provided by the sequencing facility (Q_F) to the actual value from data (Q_D)?
2. Do the errors occur at random? In which case, increasing the depth should make the base-call more accurate.

   Or are there systematic errors? In which case, increasing the depth will not improve the accuracy.
3. Different researchers use different thresholds for base-calling:

- How dominant should a base be to become the consensus base-call? 95%? 90%? 75%? 50%?
- How should this threshold change with depth? Statistically depth (i.e. sample size) and accuracy have a non-linear relationship.
- What should the Q score threshold be for usable data? Q = 3? Q = 5? Q = 10?
- Should the threshold be set at the level of a read? Or at the level of each base?

4. How should we treat gaps?
   - Gaps are not "real" bases from a sequencer; they are only introduced when the reads are aligned to find the consensus base
   - In fact, there are two kinds of gaps:
     - Gaps occur when a sequencer skips over a real base in the input fragment – a deletion
     - Gaps also occur when a sequencer wrongly "reads" a non-existent base in the input fragment – an insertion
       - all the other reads in the alignment pick up a gap at the site of an insertion.
   - Gaps do not have a Q-score; therefore, they do not have a reliability measure
   - Should the gaps be ignored while calculating the distribution of bases at a locus?

Table 3.1: Representative alignment for consensus calling

Locus	1	2	3	4	5	6	7	8	9	10
Consensus	C	T	A	C	G	A	C	T	T	C
read 1	C	G	A	C	G	A	C	---	T	---
read 2	C	G	A	T	---	A	C	---	T	---
read 3	C	G	A	C	A	A	C	---	T	---
read 4	C	G	A	C	G	A	C	---	T	C
read 5	C	A	A	C	A	A	C	---	T	C
read 6	C	A	A	C	G	A	C	---	T	---
read 7	C	A	A	C	A	A	A	T	T	---
read 8	C	---	A	C	G	A	C	---	T	---
read 9	C	G	A	T	A	A	C	---	T	---
read 10	C	G	A	C	G	A	C	---	T	---
read 11	C	A	G	C	A	A	C	---	T	C
read 12	T	T	A	C	G	A	C	---	T	---

5. How can we assign a realistic and objective error to a base-call?

All previous strategies improve the accuracy of a base-call to a greater or lesser extent. However, we did not come across any procedure which provides an error estimate separately for each called base, or a justification for the criteria used to make a base-call. Indeed, we suggest that any justification for using a particular set of criteria to call a base is contingent upon the accuracy with which we want to do the same.

In principal, even a sequencer which has random errors as large as 33% (Q ~ 5) on any particular read-base should yield a consensus base-call with confidence level greater than 99.9% for depth N=30 final probability is ~ $1/(1 + 0.5^{N/3})$. In practice, this almost never happens because of systematic errors. Researchers have generally obtained very large depths (200-5000 per locus) to "increase accuracy". But, even this does not guarantee accuracy in the presence of systematic errors.

In this chapter we present a statistical procedure to call the consensus[1] base at a locus of an aligned block. This procedure first determines the error probability of individual bases in a read either using a reference sequence or from the data itself in an iterative manner (the latter is ongoing work). This error probability can be a function of depth and actual read quality (as against the formal quality provided by the sequencer). Finally, these quantities are put into a bayesian framework to derive the error on each base-call, which can be subject to a threshold to obtain sequences of the desired accuracy.

Section 3.2: Base-calling Algorithm

At the outset we define the terms used in this chapter:

- **Reference-set** or **Reference**: A sequence with known bases for comparison with sequencer reads. The reference-set consists of 4093 non-contiguous **Reference-Base**s across the mitochondrial fragment of interest. The reference set was used
  - to develop the base-calling algorithm
  - to estimate errors of called-bases, as a function of depth
  - then – apply the algorithm to other sequences
- **Reads**: the single molecule sequence of bases output by each pore of MinION
- **Read-base** or **Reference-Locus**: Bases in reads obtained from Nanopore/MinION
- **Depth**: Number of read bases at a locus
  - Median depth for our specimens ~100
- **Base-call** or **Consensus base** or **Called-base**: Consensus of read bases at a locus

The principal steps in the base-calling procedure were as follows:

1In this chapter, we use the terms consensus base and base-call interchangeably, depending on the context.

- Construct the Reference-Set
- Use the comparison between Read-base and Reference-base to calculate the actual base error (Q_D)
- Determine the fractional distribution of bases {Gap, ACGT} at any locus

  Note that we are treating the gap as a base here. This fractional distribution may be calculated after factoring in the Q_D value of each base in the locus
- For each reference base as a function of depth, determine the probability that a particular read-base … Reference-to-Read probability distribution
  - both reference- and read-base consist of any of {Gap, ACGT} … 25 combinations in all
- Using the Bayes theorem, fractional distribution of bases at the locus and reference-to-read probability (derived from a reference set of known bases) to determine the probability of each of the 5 possible consensus bases at a locus
- Apply a minimum probability threshold to call the consensus base at each locus. We applied an error threshold of $<10^{-4}$

Section 3.2.1: Reference-Set

The frog individuals we collected had not been identified, primarily due to lack of taxonomic material. We knew that they were species of the *Philautus-Raorchestes* complex of the family Rhacophoridae and very likely to be conspecific. Therefore, we generated the reference set from published mtDNA segments of 10 different species comprising 4 from the genus *Polypedates*, 4 from the genus *Rhacophorus* and 2 *Raorchestes* species from the study site. These last two species had been previously sequenced using Illumina Miseq and are not part of the MinION dataset. These 10 mtDNA sequences were aligned in Geneious (v2022.2.2) in the region covered by our sequencing campaign. We selected those loci as the reference set where all the 10 species had the same base, i.e. 100% consensus, indicating that those bases were conserved across the Rhacophoridae family, including in the *Raorchestes* species in our sample. Figure 3.1 shows the reference set (100% consensus – individual sequence bases in grey) as well as the loci which were excluded (variable regions – individual sequence bases in colour). A total of 4093 out of 7219 bases served as the reference set.

All MinION mtDNA reads were mapped to the consensus obtained from this 10-species alignment.

Section 3.2.2: Q scores

Figure 3.2 shows the distribution of Q_F scores provided by MinION. It includes data from over 1.3 million read-bases.

1. The distributions are very different for the 4 bases. In particular, **G** distribution peaks at very small values (Q_F = 3) while T has a broader distribution with a peak around Q_F = 23. The median Q_F value for G is 11 while that for T is 19. The corresponding error probabilities are 7.9% (for G) and 1.3% (for T) – a factor of 6 difference!
2. Setting an arbitrary threshold of 6 or 8 would eliminate most of the G bases at a locus, resulting in an erroneous base-call. For instance, in Table 3.1 at locus 2: if most of the G bases were excluded from the analysis because they fell under the threshold they may actually turn out to be fewer than the A bases – which may not be of much higher quality but lie just above the threshold.

Subsequently, we calculated the actual Q_D value in the following manner: Calculate the fraction of read-bases corresponding to the reference set which are not the same as the reference-base. Calculate this fraction separately for each of {ACGT} and separately for each Q_F value. That is,

- at each of the 4093 reference loci
- count the number of read-base = A and Q_F = 13 in which the reference-base is A … z1
- count the number of read-base = A and Q_F = 13 in which the reference-base is not A … z2
- $Q_D(A, 13) = -10 \log[z2/(z1+z2)]$
- repeat this for all values of Q_F (35 values: 1-35) and for all 4 bases {ACGT} … 140 values

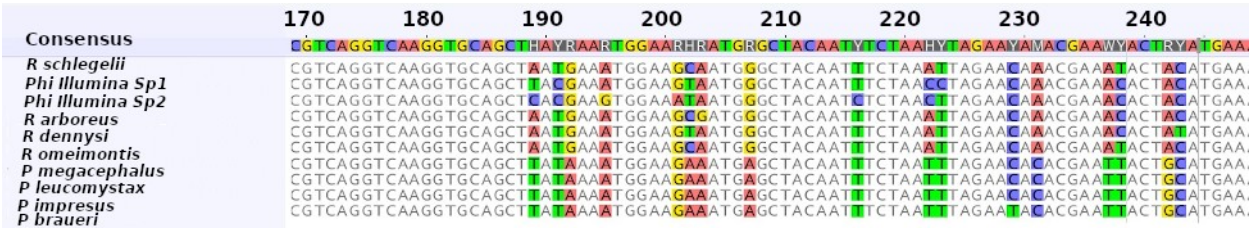


Figure 3.1: Aligned set of mtDNA from 10 different species (names on the left side of the figure) for generating the reference set, The coloured bases indicate locus in which one or more bases differ from the rest. The reference set was defined by loci with 100% consensus

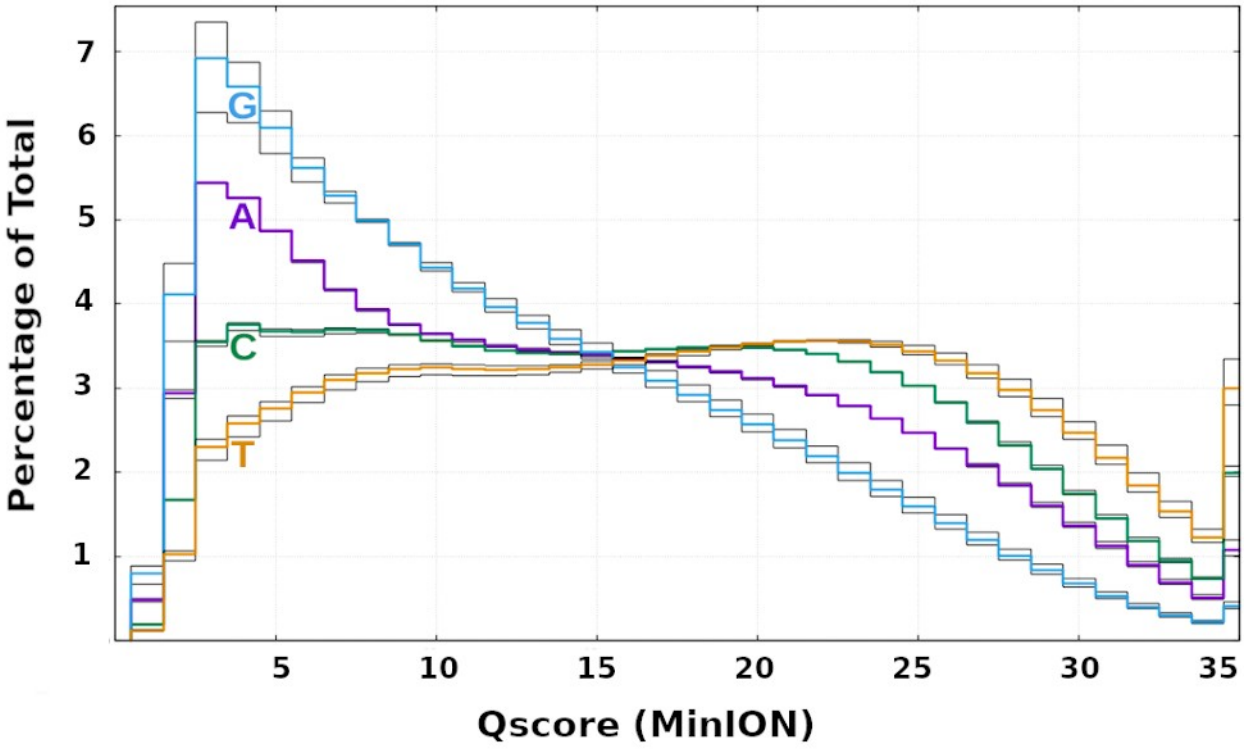


Figure 3.2: Distribution of formal Phred-Q scores (error probabilities) of ACGT bases from the fastq files provided by MinION.

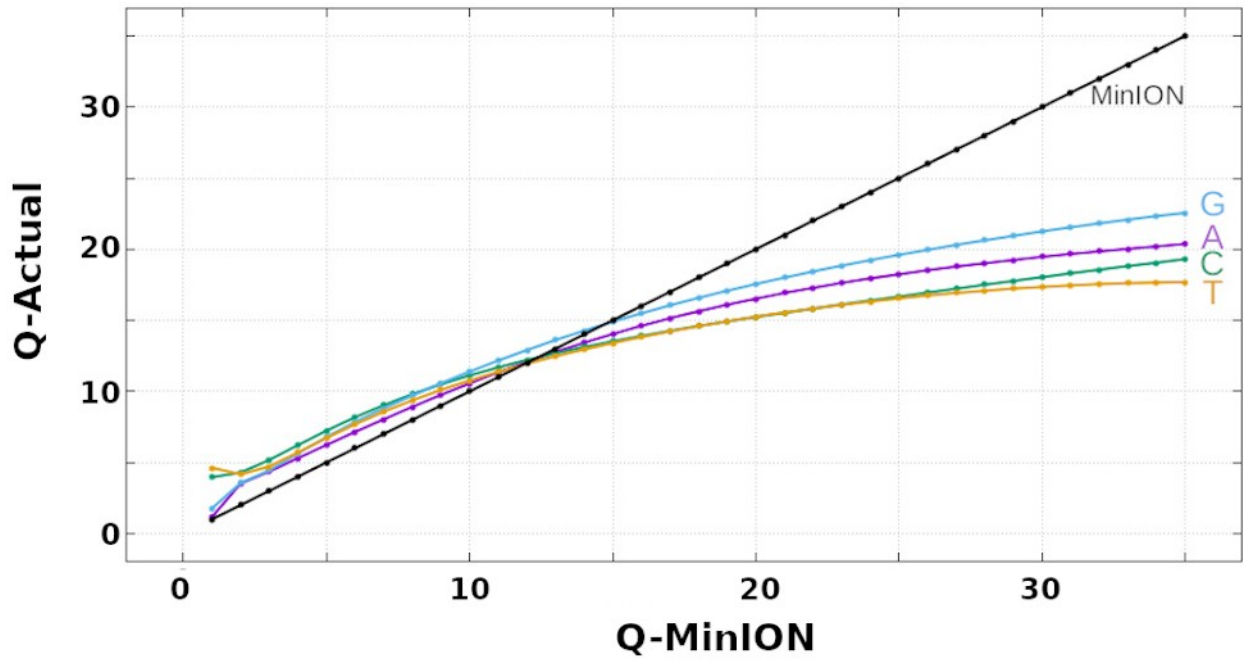


Figure 3.3: Comparison of actual calculated from data) and formal Phred Q (from MinION) for each base. Phred-Q is a measure of the error probability of a read-base.

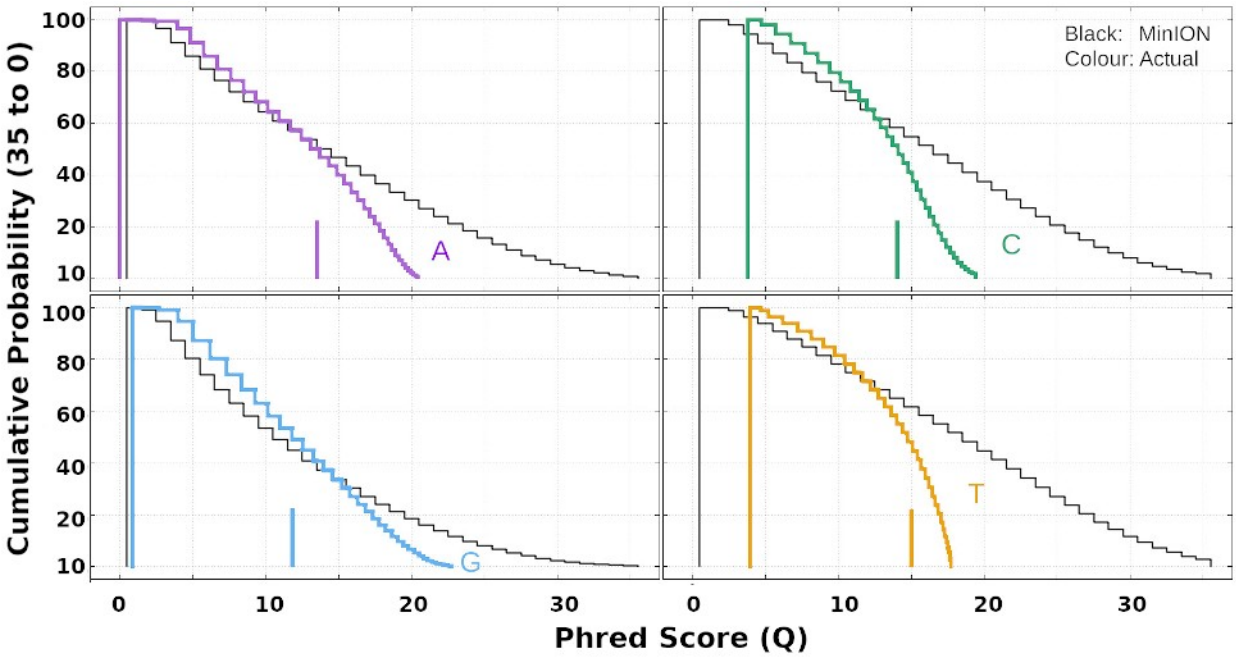


Figure 3.4: Cumulative probability distributions of of Q_F (MinION: black) and Q_D (Actual: coloured). The cumulative probability at any value q is calculated from the fraction of bases with Q > q. The coloured vertical line on X-axis is the median Q_D value of each base.

Figure 3.3 shows a plot of the comparison between Q_F and Q_D. It shows several interesting features:

1. For values of $Q_F \leq 13$ ... $Q_D > Q_F$, else $Q_D < Q_F$
2. The highest Q_D value lies between 17 (for T) and 23 (for G). It is well below the formal value which extends all the way to 40. The actual error probabilities of the "best quality" read-bases are 30 times higher than claimed by MinION

Figure 3.4 shows the cumulative probability distribution of Q_F and Q_D values separately for {ACGT}. The shapes are very different, especially at the higher end. In particular, there is a sharp fall in the distribution of T above $Q_D = 17$. The median Q_D values range from 12 to 15. Therefore, the actual errors are factors of 0.8 to 2.4 higher than the formal errors. We note that a 2.4x increase in error needs a 6x increase in depth to be compensated. Clearly, weighting the bases by Q_D while calculating the locus fraction is essential – having a single (arbitrary) threshold Q value for all bases is not appropriate for an objective recipe for base-calling.

Table 3.2: Read-bases at a locus. Q_F and Q_D are the formal (MinION) and actual data Q scores, respectively. Q-Wt is the weight of each base, being the ratio of Q_D of the read base and mean Q_D for the locus. The distribution of counts and base-fraction of the 5 bases is also listed.

Read#	Read-base	Q_F	Q_D	Q-Wt
1	A	13	12.77	1.29
2	C	5	7.24	0.73
3	A	11	11.34	1.15
4	Gap	-	---	---
5	A	14	13.42	1.36
6	A	11	11.34	1.15
7	G	5	6.79	0.69
8	A	9	9.75	0.99
9	Gap	-	---	---
10	T	4	5.70	0.58
11	T	4	5.70	0.58
12	A	9	9.75	0.99
13	A	14	13.42	1.36
14	A	12	12.07	1.22
15	T	8	9.37	0.95

Q_D Sum	128.7
Q_D Mean	9.897

Counts	Gap	A	C	G	T
Raw	2	8	1	1	3
Q-weighted	---	9.48	0.73	0.69	2.10

Base-Fraction	Gap	A	C	G	T
Raw	0.133	0.533	0.067	0.067	0.200
Q-weighted	---	0.729	0.056	0.053	0.162

Section 3.2.3: Fractional Distribution of Read-bases at a Locus

The core data input for base-calling at a locus are depth and fractional distribution of bases at the locus. Indeed, the particular method by which one determines the base fraction distribution is at the heart of the issues with base-calling.

Table 3.2 provides an example of the distribution of read-bases at a locus with depth 15. One can define the fractional distribution in two ways:

1. Fraction of counts of each of the 5-base {Gap,ACGT}
    - equivalent to the assumption that all read-bases in the locus are equally reliable.
2. Fraction of Q_D-weighted counts of each of the 4-base {ACGT}
    - Q_D is the logarithm of error. Therefore, the sum of two Q_D values is equivalent to the multiplication of the corresponding error probabilities. Therefore, simple weighting of the bases by their Q_D values is equivalent to working with their joint error probability.

We note that gaps cannot be included in the second case since they do not have a Q score.

Section 3.2.4: Reference-to-Read probability

Using the reference set one can determine the probability that a particular reference-base will result in the measurement of a particular base-fraction at a locus. This will in general be a function of the depth at the locus. For instance, considering all loci (= N_C) for which the reference-base is C

- let N_{CA} = number of loci with fraction of read-base = A is 0.2 (any value between [0:1.0])
- let N_{CC} = number of loci with fraction of read-base = C is 0.8 (any value between [0:1.0])
- For a good sequencer, N_{CA}/N_C will be closer to 0.0 and N_{CC}/N_C will be closer to 1.0

Two actual Reference-to-Read probability profiles (from our data) are shown in Figure 3.5. The X-axis of the plot is the base-fraction of the read base – in this case the read-bases are {Gap} and {T}. The reference base in both plots is {Gap}.

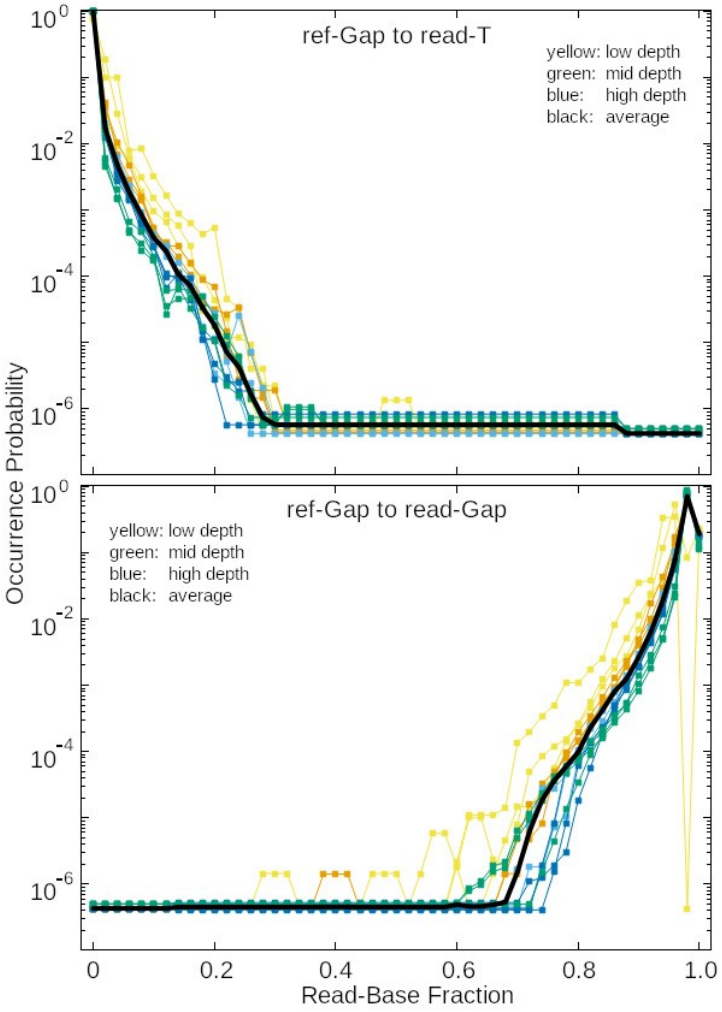


Figure 3.6: Reference-to-Read depth fraction probability profiles estimating the probability of read-base depth fraction for read-base T (upper) and Gap (lower) when reference is a Gap. The colour represent different depth classes.

The raw curves were calculated with a resolution of base-fraction = 0.02. Since the data is not continuous we then smoothed with a window size of 3 points. The raw profiles extended to about 10^{-4}. We did a linear extrapolation of the curve by another 2 orders of magnitude. This may be justified by the fact that many distributions which involve the product of probabilities of occurrence (e.g. Poisson, binomial, multinomial) have an exponential decay away from the peak. As it turns out, linear extrapolation to 10^{-6} is not really necessary – a floor at

10^{-4} or even 10^{-3} is sufficient for the purpose of estimating consensus-base error probabilities at the level of 10^{-4}.

As expected, the ref{Gap}-read{Gap} profile is non-zero mostly at values of gap-fraction closer to 1.0 … i.e. when the reference-base is {Gap}, the most probable read-base is also {Gap}. On the other hand, the ref{Gap}-read{T} profile is non-zero most at values closer to 0.0 … i.e. when the reference-base is {Gap}, the probability that the read-base is {T} is small.

The colours indicate depth – yellows: low, greens: mid and blue: high. The black is the mean curve which has also been smoothed. Clearly, the higher depth curves are also narrower near the peak – their values lie closer to the expected value (0.0 or 1.0), i.e. they are more accurate. It is reassuring that the profiles for different depths are similar.

Ref{Gap}-Read{Gap} corresponds to the case of an insertion in a sequencer. When one or a very few reads pick up {ACGT} at a locus, all the other reads in the alignment will pick up a {Gap}, i.e. the fraction of gaps at that locus will be high. On the other hand any profile of the type Ref{ACGT}-Read{Gap} corresponds to a deletion. In this case the fraction of gaps at that locus will be high. A good sequencer will only occasionally skip a base and introduce a gap at that locus.

We generated 25 such probability profiles for base-fraction from raw counts, i.e. for each combination of 5 reference base {Gap,ACGT} x 5 read-base {Gap,ACGT}
We generated 16 such probability profiles for base-fraction from Q-weighted counts, i.e. for each combination of 4 reference base {ACGT} x 4 read-base {ACGT}

We created look-up tables of the probability profiles parametrised by reference-base, read-base, 8 classes of depth, 51 classes of base-fraction. Loci with depth values in-between the nominal values of the depth classes were processed using interpolation.

Section 3.2.5: Calling the Consensus-Base

We used the reference set to calculate the probability of getting a particular base-fraction of a particular read-base … given a reference-base. Whereas, the problem we need to solve is the

inverse: given a particular observed set of values of the base-fractions of all the read-bases what is the probability of a particular consensus-base?

The Bayes theorem is the natural tool to achieve this inversion:

$$\mathbf{P(L_K \mid M) = P(M \mid L_K) \,.\, P(L_K) \,/\, P(M) \;\ldots\; \equiv p_K}$$

where,

K	:	Index denoting a particular base; K=0: Gap, 1:A, 2:C, 3:G, 4:T
M	:	measurement set, i.e, the set of all 5 base-fractions at a locus.
L_K	:	condition that the consensus base is K
$P(M \mid L_K)$	:	conditional probability of observing a particular M, given a particular reference-base K … the reference-to-read probability profiles we have previously described and calculated from the reference-set and data
$P(L_K \mid M) \equiv p_K$	:	Conditional probability that the consensus base is K given the observed measurement set … the goal of this exercise
$P(L_K)$	:	Prior probability of L_K. One can start with the distribution of {Gap,ACGT} in the reference set as a measure of $P(L_K)$ and if necessary iteratively update the values after each round.
P(M)	:	Summation of $[P(M \mid L_K) \,.\, P(L_K)]$ over all values of K

Calculate p_K for all values of K at every locus, i.e. $p_0 \equiv$Gap, $p_1 \equiv$A, $p_2 \equiv$C, $p_3 \equiv$G, $p_4 \equiv$T.

Determine the consensus-base as per a defined set of criteria.

We applied the following criteria:

- calculate $p_{4,K}$ … using Q-weighted counts for the 4-base set {ACGT}
- calculate $p_{5,K}$ … using Q-weighted counts for the 5-base set {Gap,ACGT}

- consensus base = Gap if $p_{5,0} > 0.9999$, *else*
- consensus base = B if $p_{5,B} > 0.9999$, where B is one the 4-base {ACGT}, *else*
- consensus base = B if $p_{5,B}$ is highest and $p_{5,0}$ is not second highest and $p_{4,B} > 0.9999$
  - B was the dominant base using the 5-base set but its probability was less than threshold
  - the Gap was not the second most dominant base … therefore ignore it.
  - Since Gap is not in the reckoning, use the 4-base set probability for the consensus call

  else
- etc … as per user accuracy requirement … *else*
- flag the locus by setting consensus base = N (could be any)

We followed the simple prescription of … "*when in doubt, flag the data*". Flagging F loci out of a total of T loci reduces statistical significance by sqrt{T/(T-F)}. For T ~ 10^4, erroneously flagging even F = 100 valid loci, would result in a loss in signal-to-noise by a factor of 1.005; or equivalent to working with a fragment of length 9900 instead of 10000 – quite negligible. However, wrongly including an incorrectly identified locus would greatly increase the error. Our goal is to reduce errors to 1 base in 10^4. Including just 10 incorrectly identified loci will increase the error rate by factor of 10 which will reduce phylogeny time resolution by factor of 10. Essentially, losing some data is less harmful than introducing wrong data.

Table 3.3: Base-calling results.
Good-N4: consensus {ACGT} = reference;
Bad-N4 consensus {ACGT} ≠ reference;
Flag: flagged by the procedure

10 kb fragment			
Depth	**Good-N4**	**BAD-N4**	**Flag**
38	9937	8.6	54.1
60	9949	6.2	45.3
82	9956	3.2	40.5
106	9958	3.7	38.7
132	9957	4.2	38.5
369	9960	5.6	34.8

Table 3.3 summarises the results of consensus base-calling algorithm, normalised to a 10 kb DNA sequence. "Good-N4" column represents those loci where the consensus called-base matches the reference set. "BAD-N4" column is the number of loci where called consensus base differs from that in the reference despite having error probability < 10^{-4}. The "Flag" column represents those loci where the consensus call did not meet the accuracy threshold. A fraction of these would have been part of Bad-N4 column, doubling the numbers there, thereby reducing the time resolution in a phylogenetic tree, perhaps by the same factor.

We draw attention to the fact that the values in "BAD-N4" initially decreases with depth but increases again. We do not understand this yet but it is related to the systematic effects. It may be related to the fact that as the depth increases the consensus alignment greatly increases the length of the alignment. For instance, for a depth of about 20, the alignment of the ~7000 mtDNA amplicon increased to about 10000 with the insertion of gaps; for a depth of 1000 the aligned length was about 30000. This is because a sequencer insertion in even a single read will add a gap to the alignment, and the probability of insertion at a locus increases with depth. The difficulty in accurately identifying all the spurious gaps – 23000 gaps in an alignment of length 30000 – especially in the presence of systematic effects, may be responsible for the marginal increase in errors.

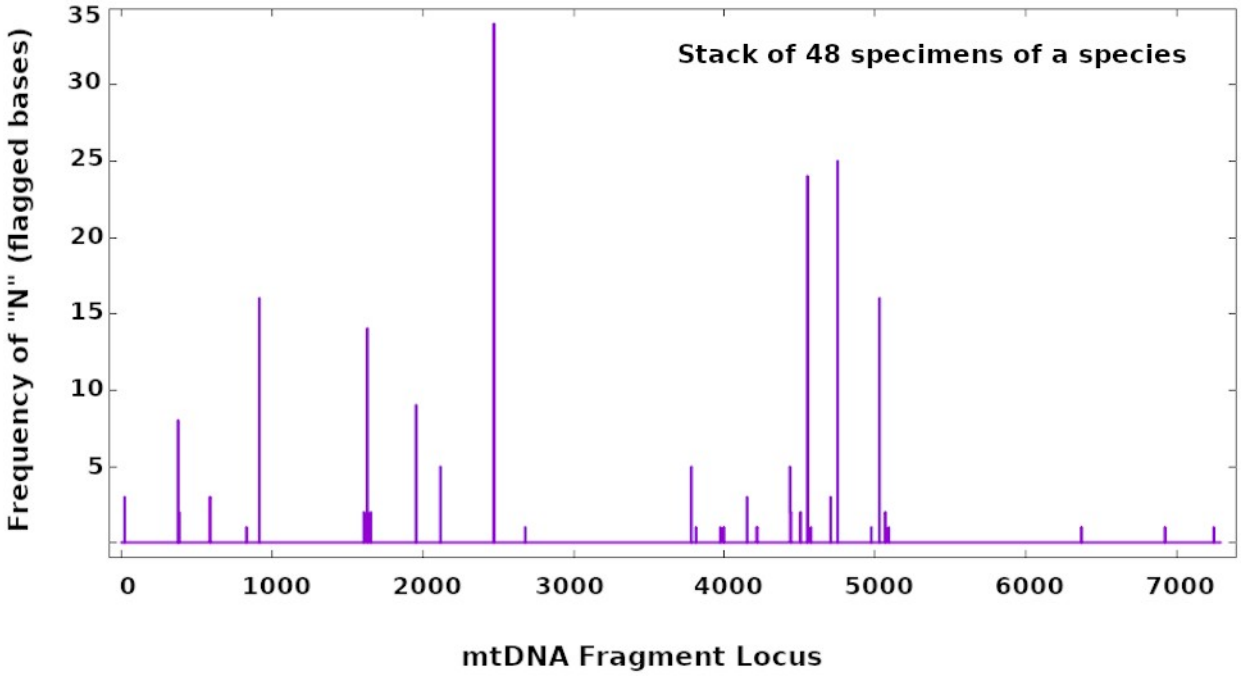


Figure 3.7: Distribution of flagged consensus bases (set to N) in an alignment of mtDNA amplicon of 48 individuals of a frog species.

This brings us to the major advantage of this technique, and one which is not reflected in the table – its ability to deal with gaps in an objective manner. Consider two samples: one with a read depth of 30, of which 27 read-bases are {ACGT}, and another with a read depth of 100 of which 73 are gaps and 27 are {ACGT}. All researchers will routinely call the consensus-base from {ACGT} in the first case. However, should the consensus-base in the second case be {Gap} or should it be called from the 27 read-bases of {ACGT}? We did not find any objective way of making this decision in published literature; our recipe provides a way. Table 3.3 deals with accuracy in the estimation of ref-base {ACGT}. Quantifying the efficacy of the procedure in identifying gaps is less straight-forward.

This method of base-calling resulted in an unexpected bonus after we aligned the mtDNA sequences from multiple individuals of one species. Figure 3.7 shows the distribution of flagged bases, i.e. those which were marked N by our algorithm because the error probability was not less than 10^{-4}. Remarkably, these flagged loci are not randomly located across the amplicon. They occur in ~30 loci across the 7200 bp amplicon but 14 loci have 3 or more flagged bases. Clearly, MinION has systematic issues associated with the sequence motifs at these loci. Base-calling at these loci without an objective error probability-based criteria will very likely result in correlated errors in the

sequences – resulting in well supported but spurious branches in the phylogenetic tree. A simple solution is to not use these loci in analysis – and our technique shows the way to identify these problematic loci.

Section 3.3: Discussion

Recent advancements in single molecule real-time (SMRT) sequencing has made it a very attractive sequencing strategy especially for projects where a reliable, time-tested reference sequence is not available. In particular, the Nanopore MinION platform offers the prospect of getting ~20 GB data in a very convenient manner and at modest cost. However, it has a reputation for poor accuracy.

The consensus base-calling technique that we have presented here solves this problem to some extent and is of particular utility to the study of non-model organisms. Our approach was tailored to suit both the scientific goals as well as financial constraints. Whole genome sequencing of 200 frog specimens was entirely out of the question; indeed, even ddRAD would have cost close to INR 4 million (USD ~40000). Our wet lab cost for this study was about INR 0.3 million.

Of course, ddRAD would have given us a lot more data, but that quantum of data was not needed, and that too for an order of magnitude increase in costs. Similarly, putting together 7000 bp of data using Sanger sequencing requires a much higher quantum of effort and at least 5 times in expense. Our approach of long-range PCR followed by Nanopore sequencing and an objective base-calling algorithm occupies a niche in between ddRAD-NGS on the one hand and Sanger on the other hand in terms of expense and effort. This will be particularly useful for intraspecific studies which requires large amounts of DNA to compensate for the short time of divergence.

The cost for increasing the size of the sequence data is merely cost of 4 or 5 more long-range PCR – about INR 300 (USD 3). The sequencing cost itself does not increase:

> the 15 GB expected from a MinION has enough capacity to provide 50 kbp sequence at an average depth of 200 for 1500 samples.

But even from a scientific perspective (of course, apart from expense – which is very much an integral part of scientific strategy!) this technique serves a particular niche. Sanger sequences of single genes are too small for accurate phylogenetic tree construction. More accurate trees using

~10 kbp can only be constructed by sequencing a dozen or more genes. Similarly, ddRAD builds up its data from small ~150 bp from across the genome. They both provide an "average" view of the phylogenetic history of the organism. With ~10 kb sequence length per gene our technique can generate accurate single-gene phylogenetic trees to understand the differing histories of genes in an organism.

Chapter 4

Intraspecific Diversity along Elevational Gradients

4.1 Introduction

The variation of species-level diversity along elevational gradients is well documented (Whittaker & Niering 1975, Terborgh 1977, Rahbek 1995, Stevens 1992, Lomolino 2001, Colwell et al. 2004, McCain & Grytnes 2010, Steinbauer et al. 2016). The observed elevational patterns are more varied than along latitudinal gradients (e.g. Miraldo et al. 2016). The observed patterns, which vary with taxa and locales, include monotonic decrease of diversity with elevation (Rahbek 1995,Vázquez & Givnish 1998, Heaney 2001, Lomolino 2001, Tang & Fang 2004, Gebrehiwot et al 2016), unimodal/peaked (Terborgh 1977, Sanders et al 2003, Bhattarai & Vetaas 2006, Acharya et al., 2011; Guo et al. 2013, Zhou et al. 2019), and low-elevation plateau followed by a decline (McCain 2009, Paudel et al. 2018, Els et al. 2021, Snider et al. 2024). These patterns are likely to be influenced by a combination of climatic, ecological and anthropogenic factors (Whittaker 1967, Gaston 2000, Brown 2001). Studies on biodiversity patterns along elevational gradients have reported mid-elevation peak for birds (Pandey et al. 2020), decline in species richness with elevation for mammals (Snider et al. 2024) and increase in species richness with elevation owing to lower anthropogenic disturbance in high elevations for plants (da Silva et al. 2014). Elevational gradients have also been reported to exert variation in diversity patterns between exotic and native plant species with the native species exhibiting high phylogenetic diversity and impacted more by ecological filtering and the exotic species exhibiting high phylogenetic clustering and impacted more by climate change (Manish Kumar, 2021).

Understanding patterns of biodiversity and their variation is therefore a key area of investigation in ecology. Elevational gradients in comparison to latitudinal gradients, offer an attractive ecological setting for investigating diversity patterns by hosting large variation in environmental conditions within a relatively short geographical space. This compactness makes the logistics of collecting data easier in some ways and especially in the case of a single slope, provides a context in which dispersal barriers have little role in determining species composition. Also, the climatic regime (e.g. dominant wind direction and rainy season), biogeographical history and the regional species pool

are largely the same across the environmental gradient. Therefore, elevational gradients are excellent for studying the impact of the environment on diversity.

The selection pressure that the environment exerts on species is actually mediated by its impact on individuals. Therefore, one expects that the steep environmental gradient (large range and short distances) may lead to locally adapted populations of a species at different elevations – i.e. genetic differentiation of populations. It may be noted that in our study site a transect of 10 km ground distance spans 3000 m in elevation and as much temperature range as latitudinal band spanning 3000 km (i.e. the length of India). Furthermore, the very short distances between the locally adapted populations may result in mixing which increases genetic diversity at any elevation.

Therefore, elevational transects are excellent contexts for studying the dynamics of intraspecific diversity[2] through the competing processes of local adaptation and impact of mixing of intraspecific diversity in the mountains. Intraspecific diversity can also be an indicator of greater prospect of survival of a species or population from a conservation perspective (e.g. Piaggio et al. 2009; Massad et al. 2024).

The elevational influence on genetic diversity can be posed at two taxonomic levels:

i. At the species level: Does the (mean) elevation at which a species is found influence its genetic diversity? This can be further explored by sampling all the elevational populations of the species, or by sampling the diversity at a single elevation.
ii. At the population level: Do the different elevational populations of a species have different levels of genetic diversity?

Section 4.1.1: Comparison of genetic diversity across elevational populations

In a meta-analysis comprising 41 studies, Ohsawa & Ide (2008) found four different patterns in plants: 26% reported the highest intraspecific diversity in the mid-elevation populations (perhaps due to optimal conditions at the centre, and suboptimal periphery), 19% showed the lowest diversity in the highest elevation populations (perhaps due to harsher conditions resulting in smaller population size, drift and short time since the last glacial maximum), 24% exhibited an increase in diversity with elevation (perhaps the high elevations acted as long term refugia with little

2 Unless otherwise mentioned intraspecific diversity and differentiation in this chapter refers to genetic diversity and differentiation – typically estimated by quantifying variation at neutral DNA loci.

anthropogenic disturbance and habitat fragmentation) and 29% had no significant correlation between diversity and elevation.

Daco et al (2022), using microsatellite markers, found no pattern of diversity in the plant *Anthyllis vulneraria* across a 500-2500 m elevational transect 2000 m (48 km ground distance) but did detect a decline in diversity across 2500 km of latitude. However, Ju et al (2022) report a peak in diversity in the mid-elevation population in the plant *Juniperus squamata*. They ascribed this to anthropogenic stress at the lower elevations and environmental harshness at higher elevations. Hahn et al (2012) reported no change in genetic diversity with elevation in 3 plant species, while Reisch & Rosbach (2020) reported that intraspecific diversity was lowest in the alpine zone.

Meng et al (2019) reported contrasting patterns in the broad-leaved tree *Euptelea pleiospermum.* They sampled along an elevational gradient spanning 900-2200 m located at the core and the edge of its latitudinal range. They observed a considerable decline in diversity with elevation at the core but a much weaker pattern at the latitudinal edge. They also found that the diversity in the latitudinal edge was less compared to the latitudinal core populations.

Low et al (2014) reported no change in genetic diversity between 0 m to 1200 m in the black fly *Simulium tani*. Polato et al (2017) showed a decline in diversity with elevation for the higher elevation montane mayfly *Baetis bicaudatus* but not in the lower elevation species (*B. Tricaudatus*). Liao et al (2024) reported association of genetic diversity with habitat for zoo-planktons when nuclear markers were used but not with mitochondrial DNA. However, there was no elevational pattern to the diversity. Giordano et al (2007) reported a decline of genetic diversity with elevation in the long-toed salamander in north-western Montana.

In summary, there is no particular pattern of elevational variation of intraspecific diversity across elevational populations of species across multiple taxa.

Section 4.1.2: Comparison of genetic diversity in species at different elevations

Thiel-Egenter et al (2008) is among the few – the only one we encountered – in which intraspecific diversity in species from different elevations were compared. They studied 22 high-elevation plants from the Alps and Carpathians using AFLP. Curiously, they found that species with narrower elevational range in the highest vegetation belts had higher genetic diversity. They also reported that wind pollinated and dispersed species had higher diversity than self- or insect-pollinated ones.

Section 4.1.3: Genetic differentiation between elevational populations of species

Moran et al (2017) reported a weak but statistically significant genetic differentiation which increased distance (across a 250 km transect) in populations of the invasive plant *Solidago canadensis* using microsatellites and cpDNA as molecular markers. Daco et al (2022) found a gradient of differentiation across both elevation (2000 m) and latitude (2500 km) in the plant species *Anthyllis vulneraria.* They ascribed both to isolation by distance.

Shi et al (2011) studied populations of *Castanopsis eyrei*, a dominant tree in subtropical forests to investigate the effects of both isolation by distance and by elevation. They sampled 24 populations between 251 and 920 m. They sequenced 8 microsatellite loci and estimated genetic distance using F_{ST}. They reported that genetic distance was correlated to both geographic and elevational distances. They concluded that the higher rates of gene flow among sites at similar elevations was a signature of isolation by elevation (i.e. local adaptation).

Hahn et al (2012) reported considerable genetic structure in the plant *Briza media* but did not discuss any correlation with distance. However, their higher elevation populations showed a greater degree of differentiation suggesting higher barriers to dispersal. A study on European larch in the French Alps along a gradient of 1350 to 2300 m found low genetic differentiation and minimal spatial genetic structure (Nardin et al. 2015).

Rufous-collared sparrow (*Zonotrichia capensis*) population in the Andes showed elevational differentiation in mitochondrial markers but not with nuclear DNA (Cheviron & Brumfield 2008). However, even the mitochondrial DNA did not show any correlation with geographical distance.

Thus, there are examples for genetic differentiation driven by distance (at same elevation) and by the environment (at different elevations). There are also instances of differentiation without any obvious correlation with either distance or elevation. The nature of the genetic markers (nuclear or mitochondrial) also seems to make a difference. Naturally, life history traits – especially dispersal – is also expected to play a role.

Section 4.1.4: This Work

We sampled 233 individuals of *Raorchestes* frogs along the 400-2400 m transect in Eaglenest, targetting approximately 10 individuals in each elevational band of 100 m. We did not know of their species identity at the time of sampling because (i) their calls are very similar and (ii) the intraspecific variation in colour forms is similar to the difference between species. Subsequent sequencing of a short mtDNA fragment showed that they belonged to 3 species which occupied largely non-overlapping elevations (Mishra 2015). We managed to sequence long-fragments (6-7 kb) of mtDNA and/or nuDNA for 172 individuals. We constructed a phylogenetic tree of these frogs with the mtDNA to estimate both genetic differentiation and diversity from SNPs therein.

The genetic data was used to address the following questions:

1. **Do the 3 elevationally segregated *Raorchestes* species have comparable intraspecific diversity?**

   Does distance from the elevational edge, both at the top and the bottom of the slope, impact intraspecific diversity, especially under the regime of climate cycling?
2. **Are the individuals of a species at different elevations adapted to the local environment (e.g. temperature)?**

   The entire geographical distance between the upper and lower elevational edges of the species ranges is only ~5 km. The single mountain slope on which the *Raorchestes* were sampled does not have any obvious barriers to the movement of frogs. Therefore, any genetic differentiation between individuals at two different elevations is most likely to be due to local adaptation.

Section 4.2: Data and Methods

Section 4.2.1: Basic Data

Phylogenetic tree constructed using neighbor-joining (NJ), maximum likelihood and bayesian inference with *Rhacophorus schlegelli* as an outgroup returned three elevationally associated clades: the low elevation species around the settlement of Khellong (n = 48), mid-elevation species around the settlement of Sessni (n = 31), high elevation species above the settlement of Bompu (n = 93). The sequences were aligned in Geneious (v2022.2.2). NJ was implemented in Geneious, maximum likelihood implemented in IQTree (v2.3.6) and bayesian inference in MrBayes(v3.2.7). The topology at the level of clades is same for all three techniques. We note that it was sufficient for us to separate the 3 species (which was done with mtDNA); this study does not require the identity of the species.[3]

Section 4.2.2: Pair-wise Phylogenetic Distance

Both objectives required calculation of pairwise phylogenetic distances which were obtained from Geneious (v2022.2.2) which provides patristic distances in the form of branch length for each pair of frog sequence.

We estimated genetic diversity within a population or within a species by the mean branch length between individuals within the group. Differentiation was estimated by the mean branch length between all possible individual pairs between pairs of populations within a species

We tested local adaptation within a species by regression of genetic distance against elevational distance. We assessed this by plotting pairwise phylogenetic distance against pairwise elevational distance for each individual within a species. The actual ground distance across the entire elevational range of the species varied from 3 km (Bompu) to 10 km (Khellong). The median post-natal dispersal distance is 100-300 m for small bodied frogs (Smith and Green 2005) and 500-1500 m for large-bodied frogs (Dolmen and Seland 2016). Therefore, 3-10 km is a very short distance even when compared with the low vagility of frogs. Therefore, any increase of genetic distance with elevational distance was much more likely to be due to ecological differences than due to isolation by distance.

3 The 3 species are referred to by the localities nearest to them: Khellong, at low elevation; Sessni, at mid-elevation; Bompu, at high elevation

Local adaptation was also explored by analysing the genetic differentiation among population pairs in the form of F_{ST} which we obtained from Arlequin v3.5. We had sampled *Raorchestes* along a continuous transect due to which there were no discrete locations where sufficient numbers of individuals were collected and therefore could be subsequently treated as populations. So, we split up the elevational ranges of each species into bands of 100 m. We analysed as a population every such band which had at least 7 individuals. Sessni, which also had the least number of individuals had only one elevational band with sufficient individuals. So, we lowered the threshold number of individuals to 5 for Sessni.

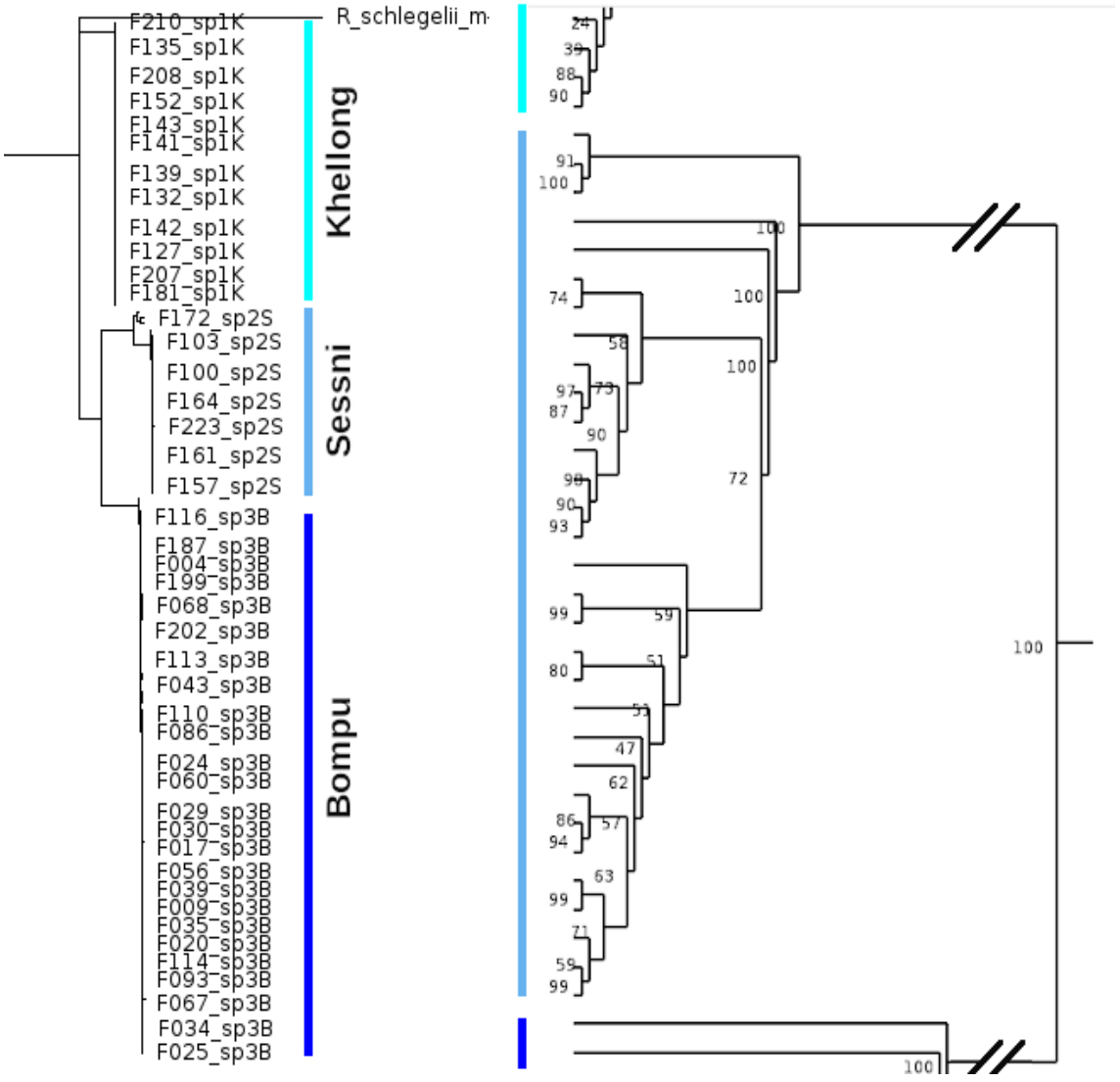


Figure 4.1: mtDNA phylogenetic tree for 172 *Raorchestes specimens (many have been excised due to limited* space) yields 3 distinct elevationally associated clades. The right half of the figure has been expanded to show the two major clades in the Sessni species.

Section 4.3: Results

Section 4.3.1: Comparison of genetic diversity in species at different elevations

Figure 4.2 and Table 4.1 show the intraspecific genetic diversity for the 3 elevationally segregated species. The middle elevational Sessni species was found to be the most diverse, as evidenced by higher mean and higher SD of its intraspecific phylogenetic distances

Table 4.1: Intraspecific pairwise phylogenetic distance for each species.

Species	N	Phylo-Dist Mean	Phylo-Dist SD	Phylo-Dist SE	Elevation mean (m)	Elevation SD (m)	Elevation SE (m)
Khellong	48	3.18×10^{-4}	2.71×10^{-4}	3.92×10^{-05}	804.4	165.28	23.9
Sessni	31	7.38×10^{-3}	1.2×10^{-2}	2.18×10^{-03}	1365.0	202.39	36.4
Bompu	93	2.44×10^{-3}	1.3×10^{-3}	1.43×10^{-04}	1966.4	200.93	20.8

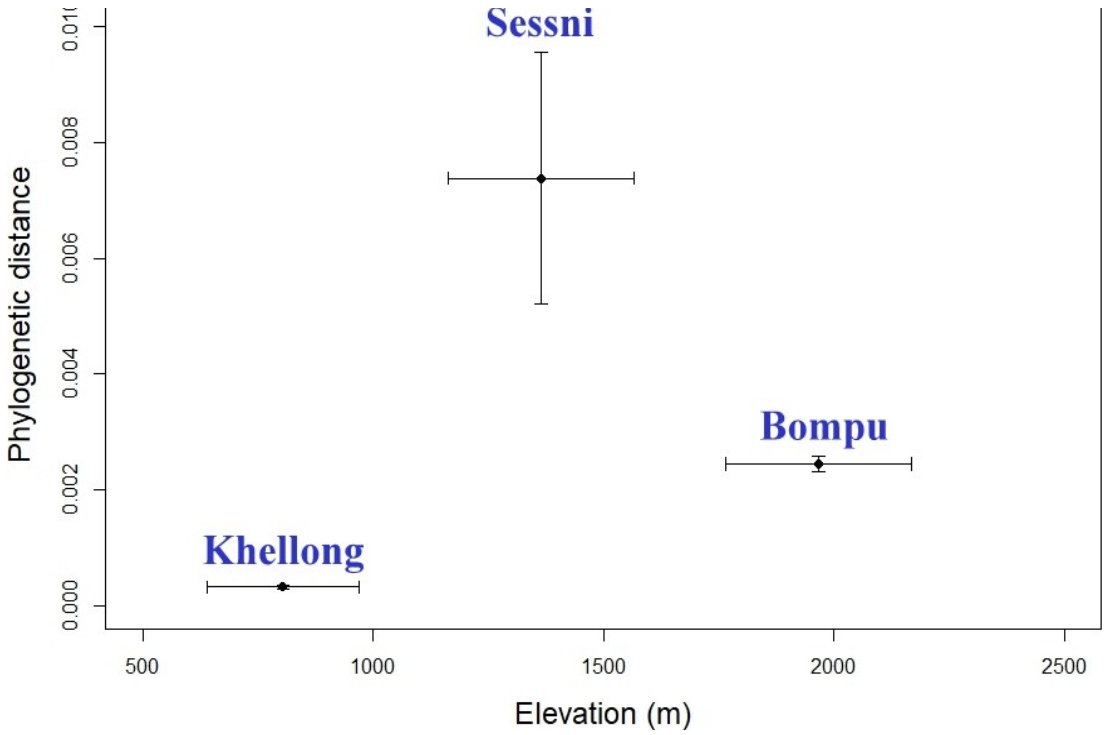


Figure 4.2: Mean and SD of pairwise intraspecific phylogenetic distance plotted against elevation for each species. Sessni has the highest intraspecific diversity. Error bars are standard deviation for phylogenetic distance and elevation (to illustrate the elevational range of the species)

Section 4.3.2: Phylogenetic distance *versus* Elevational distance within species

The results of the linear regression for pairwise phylogenetic distance and elevational distance are listed in Table 4.2 and plotted in Figure 4.3. The grey dots represent the individual pairwise phylogenetic and elevational distances. The red symbols represent the mean and SD of the individual pairwise phylogenetic distances within each bin of 50 m elevational distance. Linear regression yielded a highly significant positive relationship between phylogenetic and elevational distances for the Bompu species ($R^2 = 0.75$), but not for the Khellong and Sessni species.

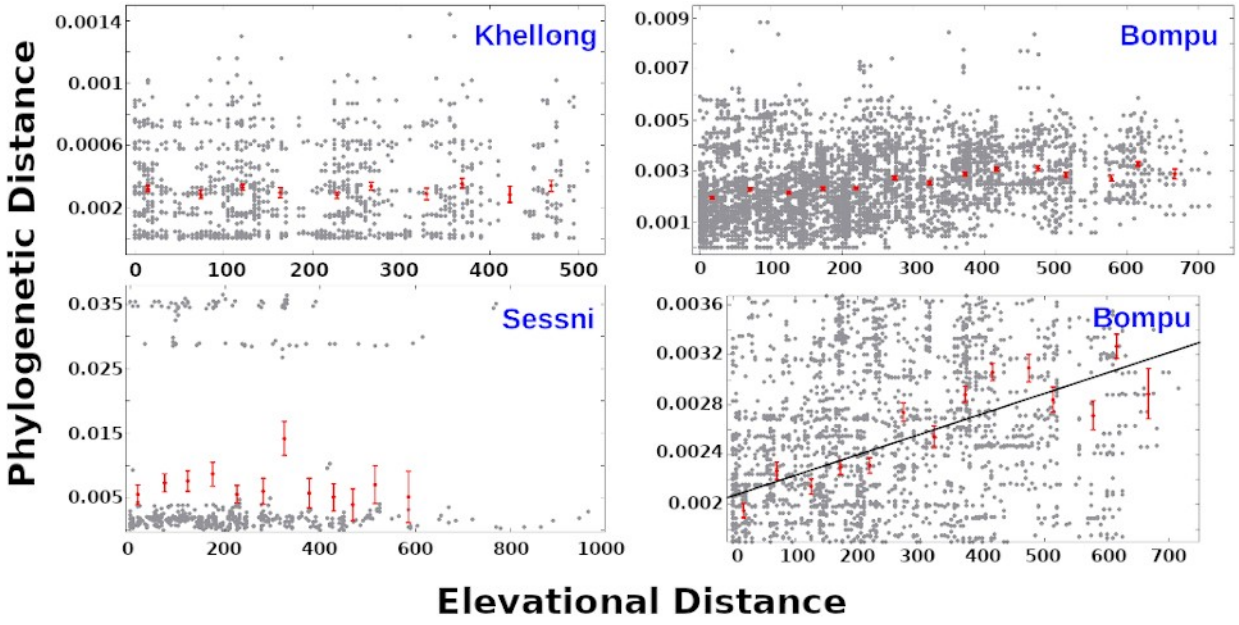


Figure 4.3: Pairwise phylogenetic distance against elevational distance. We detected a significant positive relationship for the Bompu species but not for the Khellong and Sessni species.

Table 4.2: Parameters of the linear regression between phylogenetic and elevational distances. The data is the plotted in Figure 4.3.

Species	Intercept	Slope	R^2
Khellong	$(3.06 \pm 0.18) \times 10^{-4}$	$(3.47 \pm 6.39) \times 10^{-8}$	0.04
Sessni	$(7.72 \pm 1.56) \times 10^{-3}$	$(-2.83 \pm 4.51) \times 10^{-6}$	0.04
Bompu	**$(2.07 \pm 0.11) \times 10^{-3}$**	**$(1.64 \pm 0.28) \times 10^{-6}$**	**0.75**

We also investigated population differentiation with elevational distance using F_{ST}. This analysis also shows results similar to those from phylogenetic distance, viz. a significant positive relationship between the two variables for the Bompu species but no relationship for the Khellong and Sessni species.

Table 4.3: Parameters of the linear regression to F_{ST} v/s elevational distance. Only the Bompu species shows an increasing genetic differentiation with elevation

Species	N	Intercept	Slope (10^{-4})	R^2
Khellong	6	0.0287	-0.809	0.39
Sessni	6	0.13	-1.83	0.02
Bompu	21	**-0.0394**	**4.68**	**0.59**

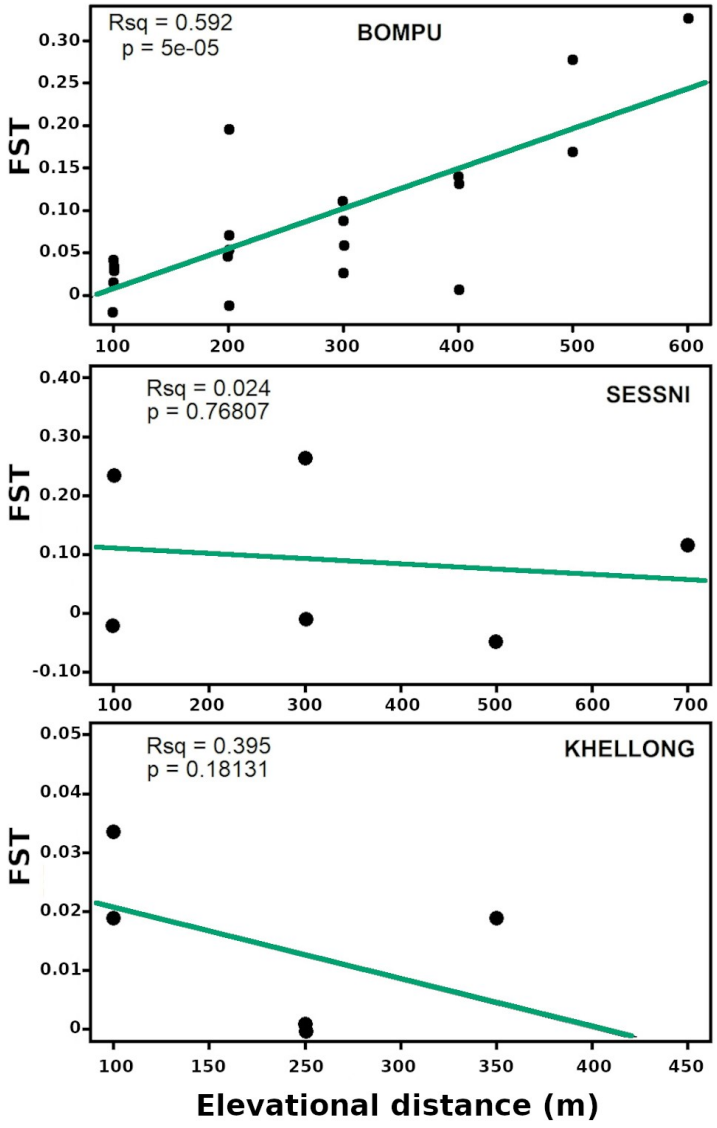


Figure 4.4: Pairwise elevational F_{ST} obtained from Arlequin plotted against elevational difference for populations of Khellong and Sessni. No significant relationship observed for either population.

Section 4.4: Discussion

Our objectives were twofold: (i) does the location of a species along the elevational gradient affect its intraspecific diversity, and (ii) is there a signature of adaptation to local conditions (using genetic differentiation as a proxy) within the elevational range of each species. We measured both diversity and differentiation using phylogenetic distance from a tree created using ~7000 bp of mtDNA.

The analysis of intraspecific diversity for the three species indicated Sessni to have the highest intraspecific diversity. We did not encounter any study which investigated a possible correlation between intraspecific diversity of a species and its location along an elevational span. The nearest that we got was Thiel-Egenter et al (2008) which investigated the intraspecific diversity within plant species as a function of their elevational range and mode of dispersal.

Frogs, being strongly tied to the climatic conditions, are expected to track their (temperature) niche during the global climatic cycles. The Khellong and Bompu species, at the lower and higher edge of the mountain may be affected by global climatic cycles. During global maxima, when temperature rises, Bompu populations would run out of space for tracking cooler temperatures higher up the mountain and during global minima, Khellong populations would run out of hill forests. The flood-plain forests of the Brahmaputra valley are quite different from the hill forests despite minimal elevational difference. Therefore, we propose that Khellong and Bompu populations get decimated during climate cycling events every 10^5 years resulting in loss of genetic diversity. However, Sessni populations located in the mid-elevation can retain their diversity better resulting in higher intraspecific variation.

The investigation into local adaptation resulted in a significant linear positive relationship between pairwise phylogenetic distance and elevational distance for only the Bompu species. The correlation is strong even though the absolute change is small ~0.16% in 1000m. This pattern was also analysed using Arlequin and Hudson's F_{ST} as a measure of pairwise population genetic differentiation. Hudson's F_{ST} (not shown here) did not show a significant relationship with elevational distance, but Arlequin's F_{ST} showed a significant positive relationship with elevational distance for the Bompu species. However, the Khellong and Sessni species did not show any such relationship with either measure of F_{ST}.

We recognise that adaptation to environmental conditions is more difficult to prove and our result only indicates a relationship between phylogenetic distance and elevational distance. However, we argue that the circumstantial evidence favours environmental adaptation over isolation-by-distance.

Firstly, the ground distance between the outermost individuals of the Bompu species is merely 2.8 km. This is not much larger than the post-natal dispersal distances of 100-300 m estimated for small bodied frogs (Alex Smith and Green, 2005). Therefore, one would expect efficient mixing of elevational populations within the timescale of tens of generations (few hundred years). Indeed we see no relationship for the Sessni and Khellong species whose ground span is 5 and 10 km (2-3 times that of the Bompu species), which we should have if isolation-by-distance operates in a similar manner in these closely related species.

Secondly, the 700 m of elevation separating the outermost Bompu species records corresponds to a temperature difference associated with 700 km of latitude (~5 ºC). This is a large enough range for, especially, temperature-sensitive taxa like frogs. It has long been postulated that higher elevations are associated with stronger environmental filters (Whittaker, 1967). Indeed, this has been shown at our field site for moths (Mungee and Athreya, 2020) and for birds (Mungee et al., 2025). This fits in with expectation of stronger environmental selection on the higher elevation species – the two lower elevation species at Khellong and Sessni do not show this relationship.

Finally, many studies have detected genetic differentiation between geographically separated populations of a species. In our case, we see a strong relationship between phylogenetic and elevational distance. This suggests a continuously and smoothly varying factor responsible for the differentiation – which suggests either isolation-by-distance or the environment. We have already argued against distance. Mungee and Athreya (2020) have shown that the composite environment consisting of temperature, precipitation, air pressure and vegetation cover changes linearly with elevation.

Hence, our contention is that the observed relationship between phylogenetic distance and elevational distance is indicative of local adaptation. Table 4.4 lists some of the elevational studies which tested for genetic differentiation as a function of geographical distance. Our detection of the relationship in the Bompu species spans the shortest distance in the list – just 2.8 km. We detected a

change of just 0.08% in 700 m. This is indeed a very small change. We could only measure it because of the novel sequencing approach that we developed, that we have described in chapters 2 and 3. This gave us ~7000 bp of mtDNA at very low cost. We are in the process of carrying out a similar analysis with nuclear DNA. Finally, we expect to settle the isolation-by-distance debate over the next few years by measuring *Raorchestes* populations along a horizontal contour well in excess of 3 km (which is the length of the elevational transect).

Table 4.4: Summary of some investigations into the genetic differentiation vs elevational distance relationship.

Source	**Taxon**	**Elev-Low (m)**	**Elev-High (m)**	**Elev-Dist (m)**	**Geog-Dist (km)**	**Correlation**
Daco et al. 2022	Plant	471	2362	1891	48	yes
Polato et al. 2017	Insect	2001	3397	1396	26	yes
Low et al. 2014	Insect	235	1405	1170	23	no
Jungels et al, 2010	Amphibian	---	---	---	60	yes
Cheviron & Brumfield, 2008	Bird	130 450	4150 4030	4020 3580	124 147	yes
This thesis		1650	2350	700	**2.8**	**yes**

This chapter closely follows a manuscript submitted to BioRxiv (https://doi.org/10.64898/2025.12.19.695388) titled *Width of Range Overlap in Congeneric Bird Species Pairs along an Elevational Gradient in the Eastern Himalayas (*authors: D. Dutta, R. Pandit, R. Athreya)

Chapter 5

Range Overlap in Congeneric Species Pairs

Section 5.1: Introduction

The range of a species is defined as the geographical extent of a species occurrence and the area occupied by populations of a species (Gaston, 1991). Species ranges offer key insights into its ecological niche and variation in response to environmental change. The borders of a species range do not extend infinitely and tend to stabilise where environmental gradient is steepest (Case and Taper, 2000) often coinciding with locations where another species is more successful – either in terms of better adaptation to abiotic (environment) and/or biotic factors (competition, predation, pathogen pressures) (Holt and Keitt, 2005). Interspecific competition and gene flow have been theoretically shown to be the principal mechanisms involving biotic factors that act synergistically to set species range limits with competition reducing local densities of a species at its range edge and gene flow from central populations reducing the frequency of locally adapted alleles (Case and Taper, 2000).

Empirical studies on factors setting range limits have focused on abiotic and biotic factors as well as the interaction between them. Factors setting range limits can possibly explain latitudinal and elevational patterns in species richness (MacArthur 1984, Rosenzweig 1996, Wiens et al. 2006; Kozak and Wiens. 2010, Giehl and Jarenkow 2012), patterns of community structure (Stephens and Wiens 2009), spread of invasive species (Peterson 2003) and allopatric speciation (Kozak and Wiens 2006). Different factors are relevant in setting range limits at different parts of a species range (Brown et al. 1996).

Role of abiotic factors in setting range limits has mostly been demonstrated with association between climatic variables and range boundaries between species. The species interactions and abiotic stress hypothesis (SIASH) predicts biotic factors to be more important in limiting species

ranges at the warm edge, i.e, at lower latitudes and elevations while abiotic factors are expected to be more important at cold edge (higher latitudes and elevation). Cahill (2014) investigated range limiting factors at the warm edge and found that abiotic factors were more important than expected under SIASH. Experimental manipulations of environmental variables ensures investigation of direct influence of abiotic factors in setting range limits. Meta-analyses of transplant experiments beyond species range have been found to be associated with decline in fitness suggesting that niche limits are indeed range limits (Hargreaves et al. 2014, Lee-Yaw et al. 2016). Abiotic factor in the form of cold water temperature affecting larval development has been demonstrated in fiddler crab *Minuca pugnax* (Sanford et. al, 2006). Expansion and range contraction following episodes of climatic warming and cooling (Crothers 1998, Parmesan, 2006) also serve as evidence for influence of abiotic factors. Abiotic factors can also structure species ranges by both inducing physiological limits as well as indirectly by affecting resource availability as demonstrated in range contraction of southern flying squirrel (Bowman et al. 2005). Abiotic factors such as temperature and ecotone (Elsen et al. 2017) was considered to be most important in setting range limits in their respective taxa and region of study.

Studies investigating the role of biotic factors alone have found them to be responsible for setting range limits in some but not all cases. Implications of range overlap has been tested in both laboratory and field settings. Arakaki et al's (2020) work involving range extending congeneric fiddler crabs did not find any evidence of competitive exclusion either in terms of habitat selection or in terms of agonistic behaviour from the species whose range recently came into overlap with a range extending congener. This implies that range overlap in these species is unlikely to threaten their future persistence. However, investigation in another taxa (birds) which were distributed along an elevational gradient found evidence for asymmetric interspecific aggression with the more aggressive species threatening the persistence of the non-aggressive congener (Jankowski, 2010; Freeman and Montgomery, 2016). Arif et al (2007) showed that the range of the dominant species was set by the environment while that of the other was limited by the dominant species. Briers (2003) inferred that the more dominant species may restrict the other to physiologically poor adapted areas where it may become susceptible to pathogen and predation pressure. This is of more concern for congeners with smaller range as any climate change mediated range extension of their more aggressive counterpart would cause them to run out of elevational space, especially at either end of the extant elevational range. Investigations into biotic factors determining range limits has

found behavioural traits such as territoriality (Freeman et al. 2019, Freeman et al. 2022) and physiological limits (Altshuler 2006) defining species ranges.

Range limits are expected to be set by an interplay of abiotic and biotic factors and some studies have actually demonstrated it. Altitude or habitat ranges of some bird and fish species are greater in mountain ranges where parapatric congeners are absent rather than in mountain ranges where congeners are present (Cadena & Loiselle 2007). Temporal variation in the position of interacting species' range limits also suggests that competitive interactions may shrink the realized range of some species to a smaller spatial area than is physiologically tolerable (Hersteinsson & Macdonald 1992). Edwards & Hernandez-Carmona (2005) demonstrated that the giant kelp, *Macrocystis pyrifera*, experienced population declines at its southern range limit owing to high temperatures and large waves during a strong El Niño event, yet was unable to recolonize formerly occupied areas after the El Niño owing to pre-emption by a competitor less affected by the abiotic change.

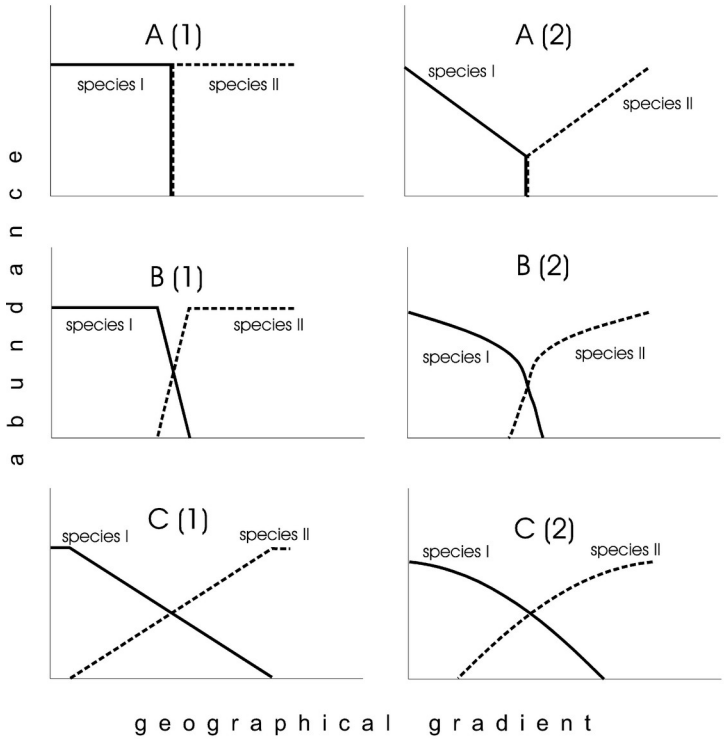


Figure 5.1: Abundance profile along environmental gradient (Muster et al. 2002)

The patterns of range boundaries arising as a result of such limiting mechanisms vary among species and have important ecological implications especially for congeneric species where range overlap is expected to induce competitive exclusion because of niche overlap. The extent of such overlaps indicate the degree of coexistence among congeners ultimately affecting their overall distribution. Range overlap patterns have also been shown to exhibit region specific variation in ecological properties of species as demonstrated in closely related species of birds whose body sizes differ more when their ranges overlap in warm as

compared to cool climates (Bothwell et al. 2015). Investigating species range patterns therefore, provides crucial insight for predicting species persistence especially when climate change mediated range shift is a major concern.

Research on climate change induced range overlap involving closely related congenerics whose ranges do not currently overlap has predicted an overall future overlap for 6.4% of new world birds, mammals and amphibians with highest for birds (11.6%) followed by mammals (4.4%) and amphibians (3.6%) (Krosby et al. 2015). Future species overlap is expected to threaten persistence through increased competition and hybridisation, leading to suboptimal adaptation for individuals at the range edge.

Muster et al (2002) have shown some of the species abundance profiles expected under the influence of both abiotic and biotic factors (see Figure 5.1). In particular, they show the impact of a second species on the range of the first. Their schematic explains the patterns of range boundaries arising as a result of interactions between various abiotic and biotic factors.

5.2 Data and Methods

5.2.1 *Raorchestes* Frogs

As described previously, the 172 individuals of *Raorchestes* frogs that we sequenced, neatly divided into 3 well-defined species-level clades, which also segregated by elevation: Khellong species (n = 48; low elevation), Sessni species (n = 31; mid elevation) and Bompu species (n = 93; high elevation). We derived the abundance ratios (lower-elevation-species to sum-of-both-species) within each 100 m elevational interval for each pair of *Raorchestes* species across their combined elevational range. We fit a logistic curve to the abundance ratio v/s elevation data of the two adjacent pairs (i.e. Khellong-Sessni and Sessni-Bompu). The sigmoid function that we fit was probability (same as abundance ratio) $p = 1/(1 + \exp(\beta_0 + \beta_1 E)$, where E is the elevation. We defined the range overlap or the range transition width as the elevational span over which the probability changed from 90% to 10%.

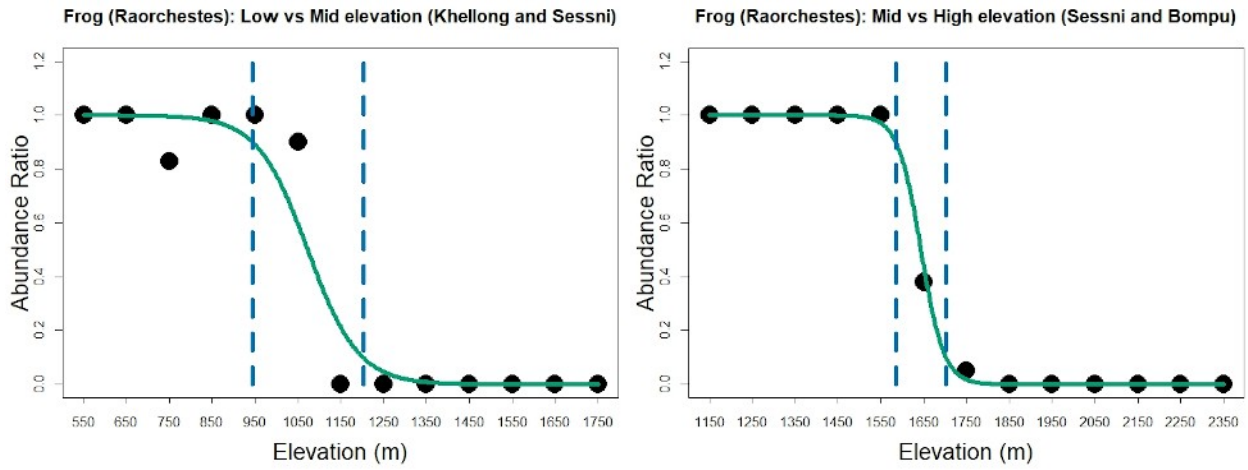


Figure 5.2: Abundance ratio of congeners modelled using a logistic regression along elevation. The solid green line is the best fit sigmoid curve. The vertical dashed lines define the region where the fraction of the lower elevation species reduces from 90% to 10% - i.e the width of the transition zone.

All analysis were done using scripts written in R (v4.3.2). The function glm in base R was used to model logistic regression.

We noted that the transition zones were quite narrow: just 117 and 258 m of elevation. Since we had only 3 species of frogs we had no way of determining if this narrowness of overlap was a general feature of either frogs or of the landscape.

However, we did have a large data base of the bird abundance distribution along the same transect. So, we used the bird data to investigate if the range overlap between congeneric species pairs was small in general. This is expected to facilitate cross-taxon comparison at a later date.

5.2.2 Bird Data

The bird data used in this study has been described elsewhere (Mungee & Athreya 2021; Mungee et al. 2025; Maitra et al. 2022). Briefly, birds were recorded on road transects of 200 m length at 49 elevations between 200 m and 2800 m. The transects were separated by 50 m in elevation. Individual birds were identified to species and recorded by a single individual (Rohan Pandit). Each transect was sampled 24 times during the summer months of May to July between 2012 and 2014. This effort yielded 15867 records of 245 species. We have followed the nomenclature and

taxonomy of Handbook of Birds of the World (HBW) and Birdlife International (http://www.birdlife.org/).

From the 245 species in the data set, we identified 107 pairs of congeneric species whose ranges either overlapped or were adjacent to each other. We fit a logistic curve to the ratio of abundance-of-lower-elevation-species to total-abundance-of-both-species within each elevational transect as a function of elevation. We grouped the 107 pairs into 5 categories, listed in Table 5.1 and illustrated in Figures 5.3-5.5:

- good fits: difference of less than 200 m or overlap in elevation between the nearest individuals of the two species and the fit had a regular sigmoid shape (Figure 5.3)
- low abundance: fewer than 10 individuals (Figure 5.4, bottom row)
- truncated: the sigmoid fit indicated that the range of one of the species extended considerably beyond the sampling limit (Figure 5.4, top row)
- poor fit: no discernible sigmoid shape to the curve (Figure 5.5, top row)
- non-contiguous: distance between the closest records of the two species more than 200 m (Figure 5.5, bottom row)

Table 5.1: Category of logistic regression fits for abundance ratios of congeneric bird species pairs

Fit category	Number
Good fit	52
Low abundance	13
Truncated	4
Poor fit	18
non-contiguous	20

We tested the "good fit" category for correlation between transition width on the one hand, and phylogenetic- and morphometric trait distances on the other. The morphometric traits included body mass, beak length, tarsus length, beak width and wing length. We expect species pairs with larger trait differences to exhibit lower interspecific competition and, therefore, have larger overlaps (see *Discussion*)

Phylogenetic distances between the bird pairs were obtained from Schumm et al (2020). The data for the 13 species missing from Schumm et al's (2020) were obtained from Mungee et al (2021) who had calculated them using Jetz et al (2012). Morphometric trait values were obtained from Schumm et al (2020). The morphometric distance was calculated as the absolute difference between lower and higher trait values of a pair taken as a percentage of the mean.

5.3 Results

The abundance ratio plot and the sigmoid fit to the data for *Raorchestes are shown in Figure 5.2.* The fit parameters are listed in Table 5.2

Table 5.2: Logistic regression model parameters for 2 pairs of *Raorchestes*

Species1	Species 2	Abundance Species 1	Abundance Species 2	Intercept	Slope	Transition Width (m)
Khellong	Sessni	48	30	18.27	– 0.02	258
Sessni	Bompu	30	93	61.48	– 0.037	117

A representational sample of the elevational profiles of the abundance ratio of bird species pairs and the logistic fits are shown in Figures 5.3-5.5. The distribution of transition widths is shown in Figure 5.6. Some of the "poor fits" are due to ecological factors, like broad ranges. However, unless otherwise mentioned all discussion will be based on the species pairs with good logistic regression.

About a quarter of the species pairs have narrow transition widths less than 100 m. However, the widths extend to considerably higher values for many species.

Figure 5.7 shows the linear regression of transition width against phylogenetic distance. Three of the species pairs, which were supposedly congenerics, turned out to have very large phylogenetic distance. The plot on the right shows the regression without those 3 data points. There is no significant correlation between transition width and phylogenetic distance.

Figure 5.8 shows the linear regression of transition width against morphometric traits. The statistics of the regressions of transition width is listed in Table 5.3. Beak width shows a correlation at the 95% confidence level. Mass and wing length show the same at a lower level of significance.

Table 5.3: Parameters of linear regression between transition width and other variables including phylogenetic distance and morphometric traits. Note that the correlation is with the percentage difference in trait between the species pair

Predictor variable	**Corr-coeff (r)**	**p-value**	
Phylogenetic Distance	0.17	0.24	
Mass	0.23	0.10	*
Beak length	0.17	0.23	
Tarsus length	0.07	0.61	
Beak width	0.30	0.03	**
Wing length	0.26	0.07	*

The correlation means that greater the difference in the trait values, the wider the overlap in the species ranges.

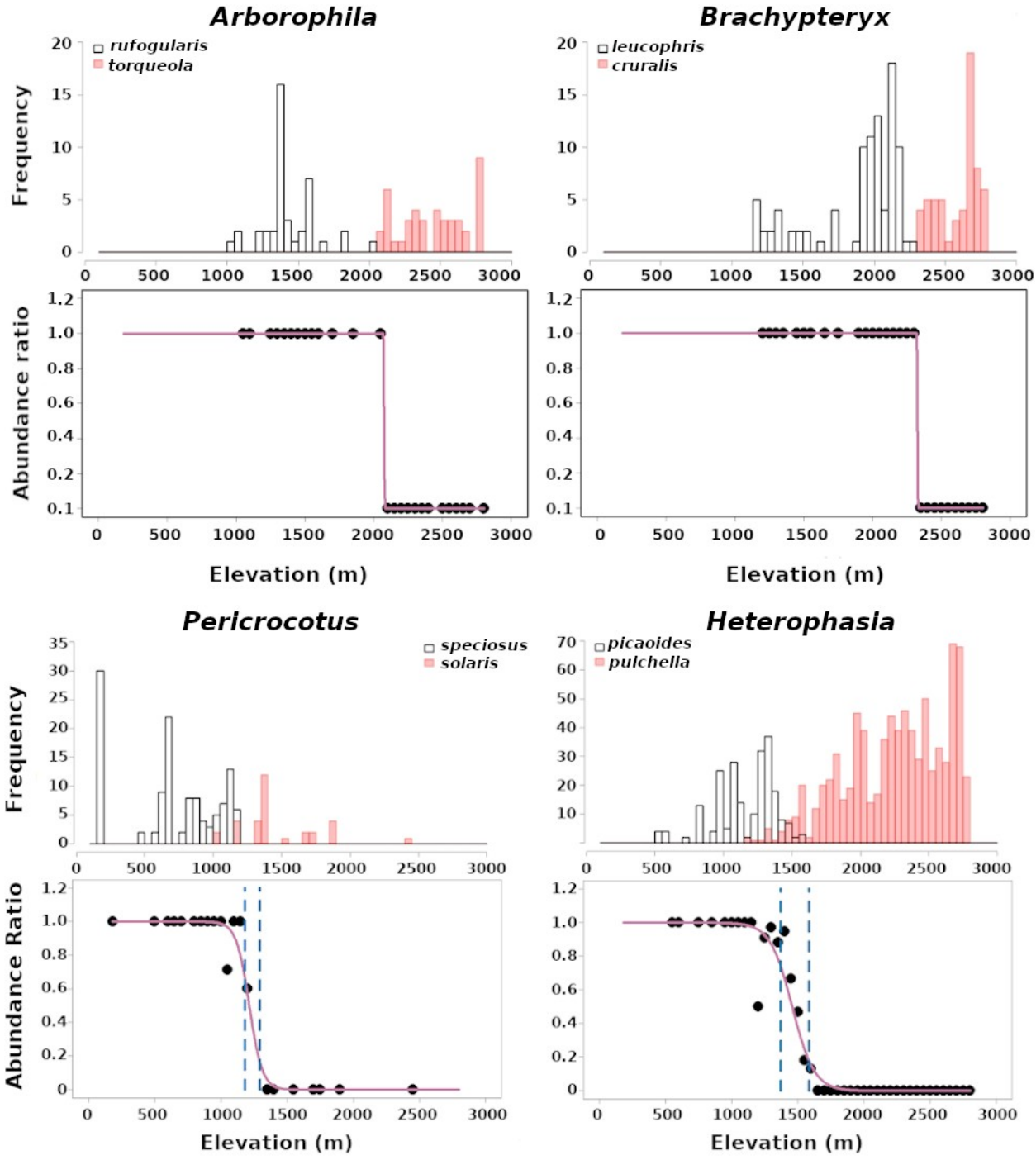


Figure 5.3: Logistic fit to abundance ratio of congeneric bird species. The transition zone, wherein the abundance ratio falls from 90% to 10% is shown by vertical dashed line. These pairs are from the "good fit" category.

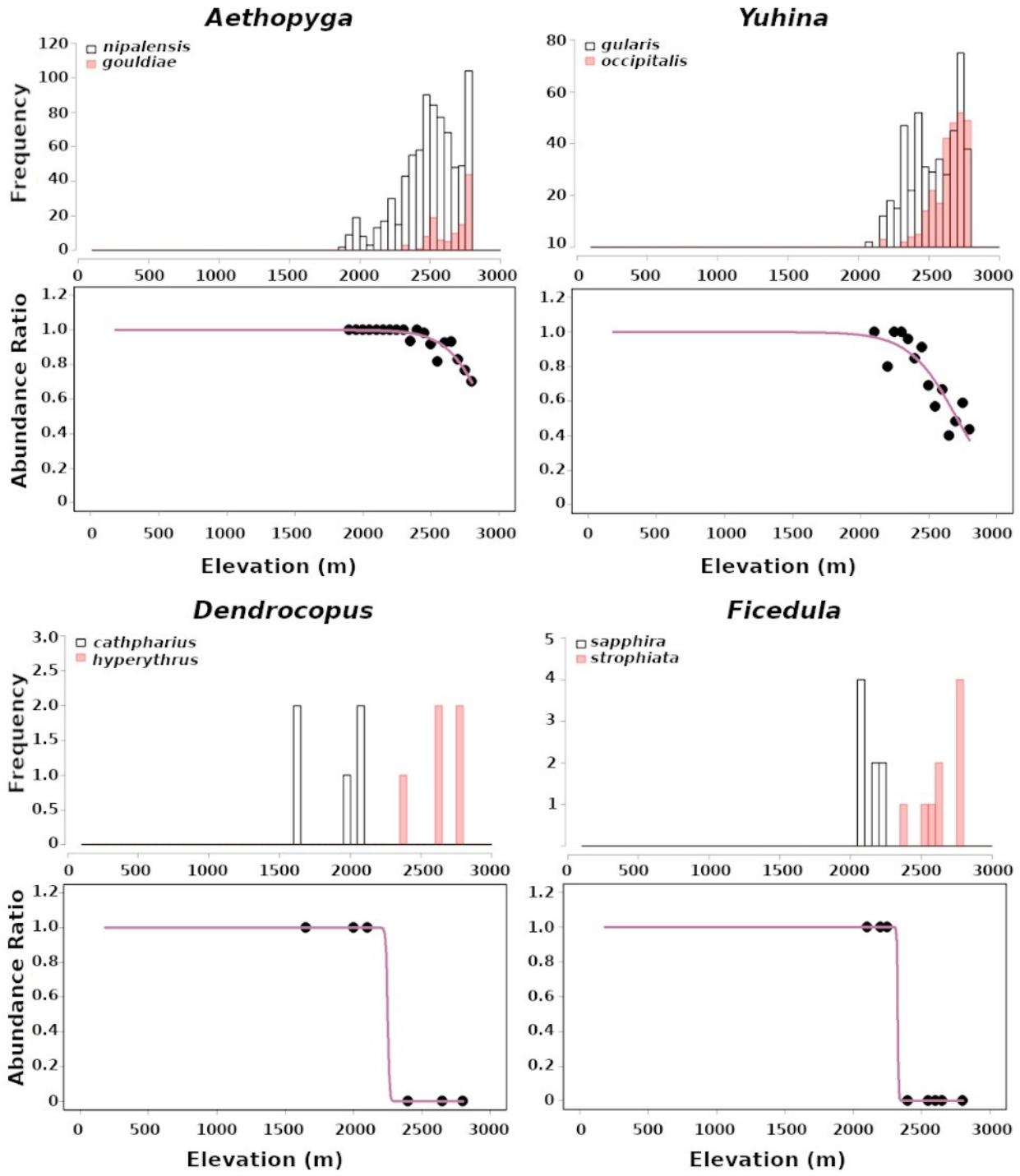


Figure 5.4: Logistic fit to abundance ratio of congeneric bird species. The upper plots are in the "truncated" category. The lower plots are "Too few" category.

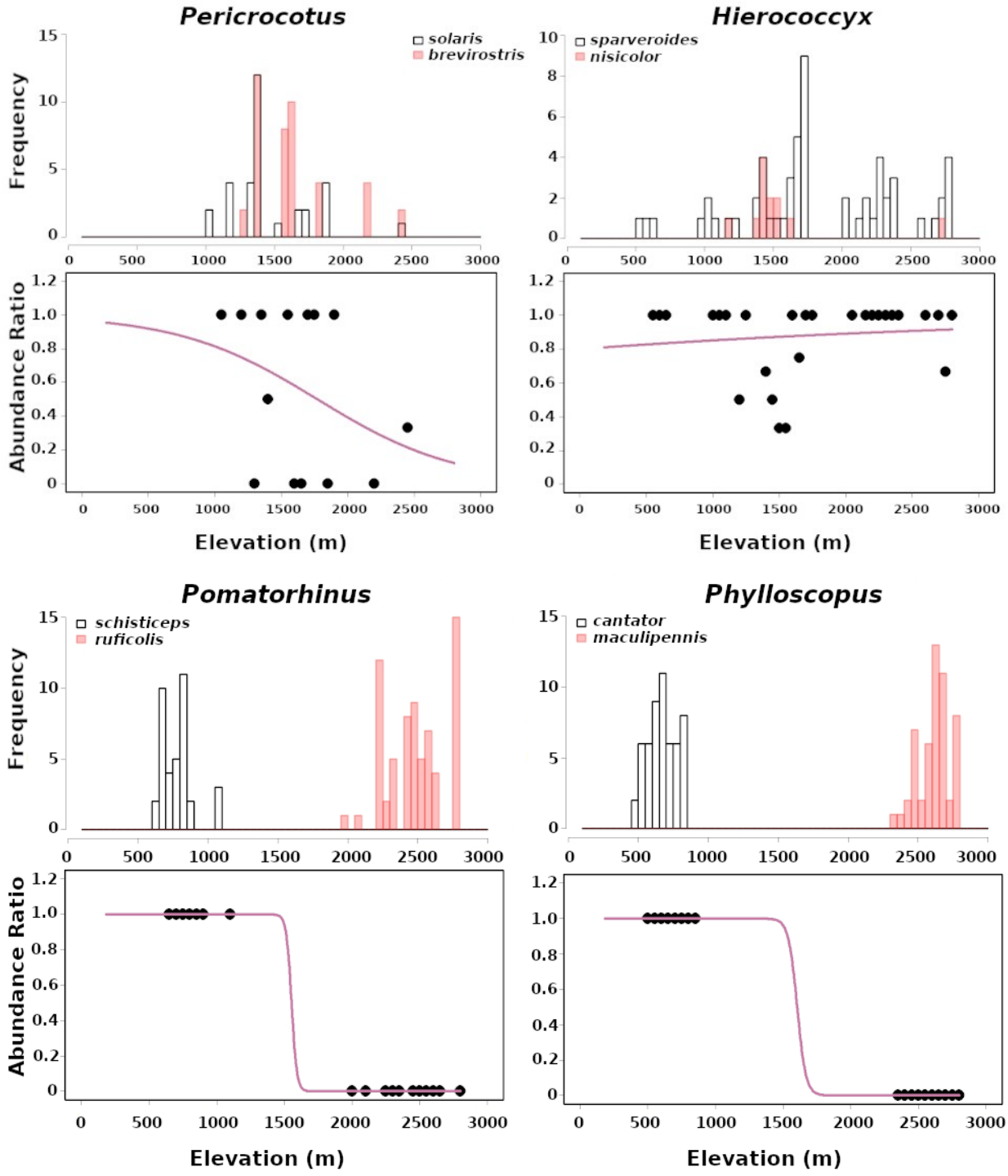


Figure 5.5: Logistic fit to abundance ratio of congeneric bird species. The upper plots are in the "poor fit" category. The lower plots are "non-contiguous" category.

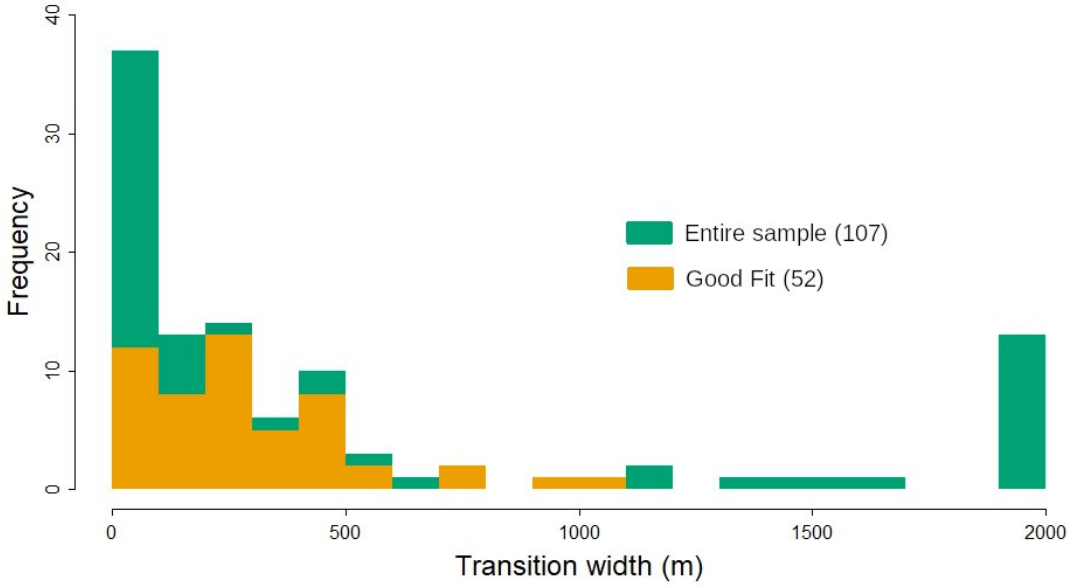


Figure 5.6: Histogram of transition width of 107 congeneric bird pair.

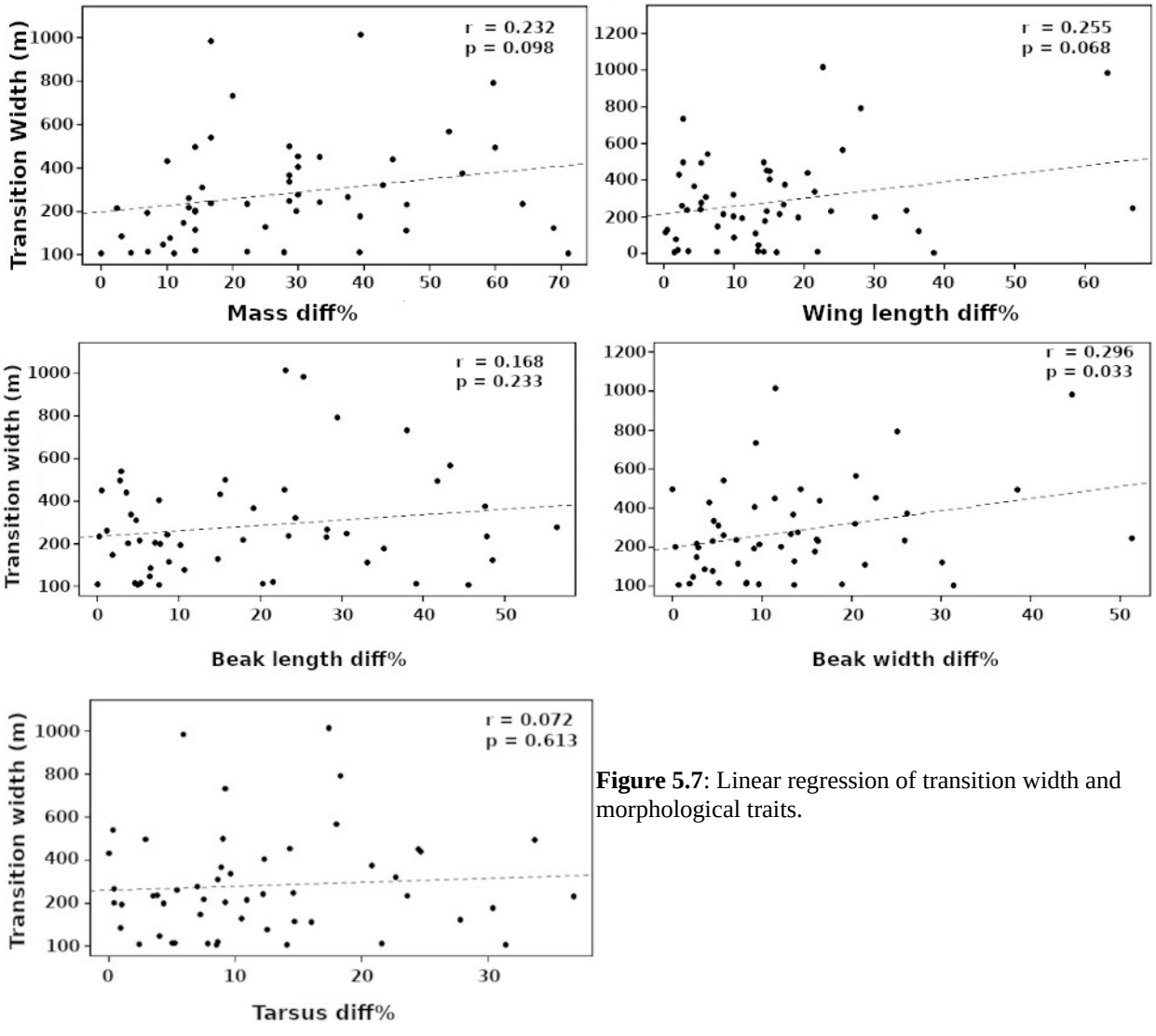


Figure 5.7: Linear regression of transition width and morphological traits.

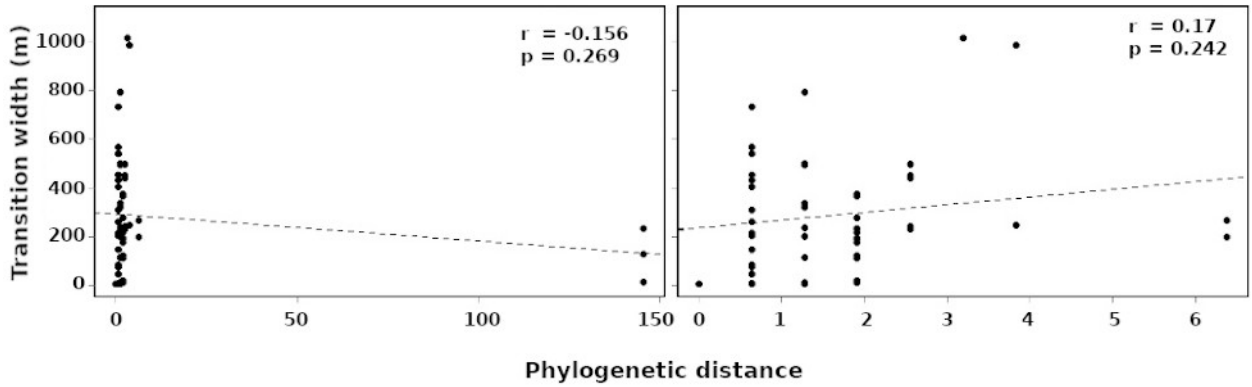


Figure 5.8: Linear regression of transition width and phylogenetic distance. Three supposedly congeneric species had very large phylogenetic distance (left). The analysis was redone without those 3 species pairs (right).

5.4 Discussion

We quantified the elevational length along which 107 congeneric pairs of birds overlap. We have labelled this elevational band as the transition zone – the region where one species gets gradually replaced by its congener. We quantified the distribution of the width of transition zones for the bird community and also investigated their relationship, if any, with phylogenetic and morpho-trait distances between the species pairs. Our analysis showed that range overlap varied from <50 m to over 500 m for bird species pairs.

Elsen et al's (2017) work in the western Himalayas showed that the distribution structure and range limits of birds found species ranges were mostly constrained by temperature followed by ecotone and competition. Srinivasan et al (2018) compared elevational distribution of bird communities between seasonal, species-poor western Himalayas and less seasonal species-rich eastern Himalayas. They found that western Himalayan bird community structure was influenced more by temperature and the eastern Himalayan one by competition. Jankowski et al (2010) found that competition was the chief driver of range limits in the low-, mid- and high-elevation nightingale thrushes in Costa Rica. The mid elevation species was found to react more aggressively to its high elevation counterpart's song which can shrink the latter's range more in the face of climate mediated range shifts. Cadena & Loiselle (2007) identified competition as well as autoecology-environment as factors impacting the ranges of Buarremon brush-finches in neotropical montane forests.

Freeman et al (2019; 2022) specifically investigated the relationship between range overlap and several other factors including interference competition, phylogenetic distance and beak divergence. They found a negative correlation with competition but not with the other two. On the other hand we do find a positive correlation with beak depth difference. We note that the definition of overlap is different in the two studies: they defined overlap as the fraction of the smaller range contained within the larger range. Whereas, we define overlap as the width of the transition zone. Whether and how this may make a difference to the results remains to be seen.

The relationship between spatial coexistence and divergent ecological traits is well established across taxa in the evolutionary ecology literature (Adams 2004, Pfennig & Pfennig 2009, Zou et al. 2021, Pigot et al. 2018). Therefore, we expected a pattern of increasing transition width with phylogenetic distance and the trait difference percentage between congeneric pairs. The results however fail to establish relationship with our predictor variables except beak width. While the absence of such relationship makes it difficult to infer any ecological or evolutionary mechanism driving such patterns, existing literature on range overlap patterns along elevational gradients indicate that such a lack of relationship is not unexpected (e.g. Freeman et al. 2019)

Evolutionary relatedness has been considered a driver for competition resulting in range exclusion (Violle et al. 2011, Davies 2006), though, its role has been challenged by others (Mayfield & Levine 2010, Godoy et al. 2014, Germain et al. 2016). Our result does not show any significant link between range overlap and degree of phylogenetic relatedness.

We found a significant positive relationship between transition width and beak width difference (but with none of the other traits tested, viz. mass, wing length, beak length and tarsus length). Studies on bird functional traits related to resource use indicate that beak width is associated with resource type and mechanistic performance whereas beak length affects foraging access (Navalón et al. 2019, Anderson & Weir 2021).

This work was motivated by the sharp transition width we saw in the two *Raorchestes* species pairs. The bird data analysis shows that the transition width extends from less than 50 m to more than 500 m. At this stage we cannot conclude if narrow transition widths is a feature of the *Raorchestes* community, or more generally of frogs. We note that these frogs are characterised by the very loud

and incessant calling by which they advertise their territories, far more than most birds. They also differ from birds in having very low mobility and in being ectotherms. Whether these differences will result influence the distribution of transition widths can only be answered by a future study.

Table 5.4: Model parameters of logistic regression fit to plots of abundance ratio of congeneric bird species pairs

Family	Genus	Species1	Species2	N1	N2	Intercept	Slope	Transition
							Prob/100 (m)	Width (m)
Campephagidae	*Pericrocotus*	*speciosus*	*brevirostris*	121	42	524	-41.8	10.5
Campephagidae	*Pericrocotus*	*speciosus*	*solaris*	121	32	49	-4.0	110.8
Cettiidae	*Abroscopus*	*albogularis*	*schisticeps*	68	194	260	-18.8	23.4
Cettiidae	*Abroscopus*	*superciliaris*	*albogularis*	33	68	7	-0.9	499.3
Cettiidae	*Abroscopus*	*superciliaris*	*schisticeps*	33	194	132	-10.4	42.4
Cettiidae	*Tesia*	*olivea*	*cyaniventer*	93	135	22	-1.1	405.5
Columbidae	*Treron*	*apicauda*	*sphenurus*	50	49	330	-43.9	10.0
Corvidae	*Dendrocitta*	*formosae*	*frontalis*	17	12	42	-5.1	86.4
Cuculidae	*Cuculus*	*micropterus*	*poliocephalus*	66	33	13	-0.6	792.8
Dicruridae	*Dicrurus*	*leucophaeus*	*aeneus*	88	234	-7	0.4	1014.8
Dicruridae	*Dicrurus*	*paradiseus*	*remifer*	14	41	21	-2.5	177.5
Leiothrichidae	*Actinodura*	*egertoni*	*waldeni*	252	87	79	-3.6	122.8
Leiothrichidae	*Actinodura2*	*cyanouroptera*	*strigula*	111	191	136	-5.7	77.1
Leiothrichidae	*Heterophasia*	*picaoides*	*pulchella*	218	816	30	-2.0	215.1
Leiothrichidae	*Liocichla*	*phoenicea*	*bugunorum*	57	14	401	-19.4	22.7
Leiothrichidae	*Pterorhinus*	*gularis*	*caerulatus*	46	73	334	-23.9	18.4
Leiothrichidae	*Pterorhinus*	*pectoralis*	*caerulatus*	92	73	129	-10.2	42.9
Leiothrichidae	*Trochalopteron*	*squamatum*	*erythrocephalum*	102	208	17	-0.9	497.7
Megalaimidae	*Psilopogon*	*asiaticus*	*franklinii*	115	94	21	-2.2	200.8
Megalaimidae	*Psilopogon*	*asiaticus*	*virens*	115	95	12	-1.2	374.6
Muscicapidae	*Brachypteryx*	*leucophris*	*cruralis*	93	60	1719	-74.0	5.9
Muscicapidae	*Niltava*	*grandis*	*sundara*	102	124	20	-1.0	452.9
Muscicapidae	*Niltava*	*macgrigoriae*	*grandis*	112	102	16	-0.9	493.9
Muscicapidae	*Niltava*	*macgrigoriae*	*sundara*	112	124	26	-1.4	320.8
Nectariniidae	*Aethopyga*	*saturata*	*gouldiae*	157	111	43	-1.9	236.9
Paradoxornithidae	*Paradoxornis2*	*atrosuperciliaris*	*nipalensis*	30	64	84	-5.5	80.2
Pellorneidae	*Napothera*	*epilepidota*	*malacoptila*	14	16	7	-0.6	733.4
Pellorneidae	*Schoeniparus*	*cinereus*	*castaneceps*	361	139	32	-1.4	309.7
Phasianidae	*Arborophila*	*mandellii*	*torqueola*	50	44	586	-29.0	15.2
Phasianidae	*Arborophila*	*rufogularis*	*mandellii*	42	50	31	-1.9	233.2
Phasianidae	*Arborophila*	*rufogularis*	*torqueola*	42	44	1649	-79.5	5.5
Phylloscopidae	*Phylloscopus*	*cantator*	*maculipennis*	54	53	59	-3.6	121.6
Phylloscopidae	*Phylloscopus*	*cantator*	*magnirostris*	54	76	59	-3.5	124.9
Phylloscopidae	*Phylloscopus*	*cantator*	*reguloides*	54	215	95	-8.1	54.0
Phylloscopidae	*Phylloscopus*	*cantator*	*xanthoschistos*	54	84	32	-3.8	115.7
Phylloscopidae	*Phylloscopus*	*reguloides*	*maculipennis*	215	53	43	-1.6	266.3
Phylloscopidae	*Phylloscopus*	*reguloides*	*magnirostris*	215	76	47	-1.8	241.3
Phylloscopidae	*Phylloscopus*	*xanthoschistos*	*maculipennis*	84	53	158	-7.7	57.1
Phylloscopidae	*Phylloscopus*	*xanthoschistos*	*magnirostris*	84	76	148	-7.1	61.8
Phylloscopidae	*Phylloscopus*	*xanthoschistos*	*reguloides*	84	215	52	-3.0	148.5
Phylloscopidae	*Seicercus*	*castaniceps*	*whistleri*	181	207	43	-1.8	246.8
Phylloscopidae	*Seicercus*	*intermedius*	*whistleri*	71	207	50	-2.2	197.7
Phylloscopidae	*Seicercus*	*poliogenys*	*castaniceps*	206	181	9	-0.4	984.0
Phylloscopidae	*Seicercus*	*poliogenys*	*whistleri*	206	207	50	-2.2	203.4
Sittidae	*Sitta*	*cinnamoventris*	*formosa*	14	20	162	-14.5	30.3
Sittidae	*Sitta*	*cinnamoventris*	*himalayensis*	14	11	65	-4.1	108.2
Sittidae	*Sitta*	*formosa*	*himalayensis*	20	11	198	-9.9	44.3
Timaliidae	*Cyanoderma*	*chrysaeum*	*ruficeps*	356	307	20	-1.0	431.1
Timaliidae	*Pomatorhinus*	*ferruginosus*	*ruficollis*	47	69	415	-21.8	20.1
Timaliidae	*Pomatorhinus*	*ferruginosus*	*superciliaris*	47	22	30	-1.6	277.5
Timaliidae	*Pomatorhinus*	*schisticeps*	*ferruginosus*	37	47	24	-2.3	191.8
Timaliidae	*Pomatorhinus*	*schisticeps*	*ruficollis*	37	69	88	-5.6	78.3

In Conclusion

We have demonstrated a hybrid strategy of DNA sequencing which can provide 10-50 kbp of data per specimen which occupies a niche in between Sanger sequencing strategy on the one hand and ddRAD on the other, in terms of scientific research, effort and expense. It is an order of magnitude less expensive than either. This strategy offers the most efficacious manner of study – especially – ecological and evolutionary processes which require the accurate measurement of intraspecific genetic diversity. In this context, evolutionary studies on the timescale of climate cycles may be targeted using this technique. Also, this may well be the best currently available tool to explore the world of single gene phylogenies to understand how they differ from the over-all genome.

While we developed the technique to estimate the error on individual base-calls with the Nanopore platform in mind, it is a general enough strategy that can be applied to any platform. The application to Nanopore was only motivated by that platform's lower accuracy.

At the moment we have only applied the technique to the mitochondrial fragment. We are currently in the process of applying the same to nuclear genes. The difference is that it is easier to find large sections of conserved regions in the mitochondria and more mitochondrial genomes are known. We expect to shortly identify the best recipe for dealing with nuclear genes.

Finally, this work was carried out in the eastern Himalayas of Arunachal Pradesh in north-east India. That region is among the most diverse in the world. With under 3% of India's land area, it hosts almost two-thirds of the country's biodiversity. Much of the area has been explored ecologically only in the last two decades, and much remains to be explored.

One of the principal causes of the diversity is almost certainly the complex connectivity of the terrain with steep mountain slopes and deep river valleys. The rich organismal diversity of the area has to be anchored to the genetic diversity at both the species and the intraspecific level. This thesis shows the road forward.

References

1. Adams, D. C. (2004). Character displacement via aggressive interference in Appalachian salamanders. Ecology, 85(10), 2664-2670.
2. Agarwal, I., Mistry, V. K., & Athreya, R. (2010). A preliminary checklist of the reptiles of Eaglenest Wildlife Sanctuary, West Kameng District, Arunachal Pradesh, India. Russian Journal of Herpetology, 17(2), 81-93.
3. Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. Nature Reviews Genetics, 11(10), 697–709
4. Allen, A. P., & Gillooly, J. F. (2006). Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. Ecology letters, 9(8), 947-954.
5. Altshuler, D. L. (2006). Flight performance and competitive displacement of hummingbirds across elevational gradients. The American Naturalist, 167(2), 216-229
6. Anderson, S. A., & Weir, J. T. (2021). Character displacement drives trait divergence in a continental fauna. Proceedings of the National Academy of Sciences, 118(20), e2021209118.
7. Arakaki, J. Y., De Grande, F. R., Arvigo, A. L., Pardo, J. C. F., Fogo, B. R., Sanches, F. H., ... & Costa, T. M. (2020). Battle of the borders: Is a range-extending fiddler crab affecting the spatial niche of a congener species?. Journal of Experimental Marine Biology and Ecology, 532, 151445.
8. Arif S, Adams DC, Wicknick JA. (2007). Bioclimatic modelling, morphology, and behavior reveal alternative mechanisms regulating the distributions of two parapatric salamander species. Evol. Ecol. Res. 9:843–54
9. Athreya, R., 2006a. Eaglenest Biodiversity Project (2003–2006): Conservation Resources for Eaglenest Wildlife Sanctuary, Kaati Trust, Pune.
10. Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., ... & Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual review of ecology and systematics, 489-522.
11. Azam, S., Thakur, V., Ruperao, P., Shah, T., Balaji, J., Amindala, B., ... & Varshney, R. K. (2012). Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results to identify SNPs in chickpea (Cicer arietinum; Fabaceae), a crop species without a reference genome. American journal of botany, 99(2), 186-192

12. B. G. Freeman, J. A. Tobias, D. Schluter, (2019). Behavior influences range limits and patterns of coexistence across an elevational gradient in tropical birds. Ecography. 42, 1832–1840
13. Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS one, 3(10), e3376.
14. Bhattarai, K. R., & Vetaas, O. R. (2006). Can Rapoport's rule explain tree species richness along the Himalayan elevation gradient, Nepal?. Diversity and distributions, 12(4), 373-378.
15. Biju, S. D., & Bossuyt, F. (2003). New frog family from India reveals an ancient biogeographical link with the Seychelles. Nature, 425(6959), 711-714.
16. Bonacum, J., DeSalle, R., O'Grady, P., Olivera, D. S. C. G., Wintermute, J., & Zilversmit, M. J. D. I. S. (2001). New nuclear and mitochondrial primers for systematics and comparative genomics in Drosophilidae. Drosophila Information Service, 84, 201-204.
17. Bonfield, J. K., & Staden, R. (1995). The application of numerical estimates of base calling accuracy to DNA sequencing projects. Nucleic Acids Research, 23(8), 1406-1410
18. Bossuyt, F., & Milinkovitch, M. C. (2000). Convergent adaptive radiations in Madagascan and Asian ranid frogs reveal covariation between larval and adult traits. Proceedings of the national Academy of Sciences, 97(12), 6585-6590.
19. Bothwell, E., Montgomerie, R., Lougheed, S. C., & Martin, P. R. (2015). Closely related species of birds differ more in body size when their ranges overlap—in warm, but not cool, climates. Evolution, 69(7), 1701-1712
20. Bowman J, Holloway GL, Malcolm JR, Middel KR, Wilson PJ. (2005). Northern range boundary dynamics of southern flying squirrels: evidence of an energetic bottleneck. Can. J. Zool. 83:1486–94
21. Briers RA. (2003). Range limits and parasite prevalence in a freshwater snail. Proc. R. Soc. London Ser. B 270:S178–80
22. Brown, J.H., Stevens, G.C. & Kaufman, D.M. (1996) The geographic range: size, shape, boundaries, and internal structure. Annual Review of Ecology and Systematics, 27, 597–623
23. Brown, J. H. (2001). Mammals on mountainsides: elevational patterns of diversity. Global Ecology and Biogeography, 10(1), 101-109.

24. Cadena CD, Loiselle BA. (2007). Limits to elevational distributions in two species of emberizine finches: disentangling the role of interspecific competition, autoecology, and geographic variation in the environment. Ecography 30:491–504
25. Cahill, A. E., M. E. Aiello-Lammens, M. Caitlin Fisher-Reid, et al. 2014. “Causes of Warm-Edge Range Limits: Systematic Review, Proximate Factors and Implications for Climate Change.” Journal of Biogeography 41: 429–442
26. Case and Taper (2000). Interspecific Competition, Environmental Gradients, GeneFlow, and the Coevolution of Species’ Borders. The American Naturalist, Vol 155, 583 – 605.
27. Campbell, K. S., Baltensperger, A. P., & Kerby, J. L. (2024). Random Frogs: using future climate and land-use scenarios to predict amphibian distribution change in the Upper Missouri River Basin. Landscape Ecology, 39(3), 61.
28. Cheviron, Z. A., & Brumfield, R. T. (2009). Migration-selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows (Zonotrichia capensis) along an elevational gradient. Evolution, 63(6), 1593-1605
29. Chen, I. C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid range shifts of species associated with high levels of climate warming. Science, 333(6045), 1024-1026.
30. Choudhury, A. (2003). Birds of Eaglenest Wildlife Sanctuary and Sessa Orchid Sanctuary , Arunachal Pradesh , India, 19, 1–13
31. Colwell, R. K., Rahbek, C., & Gotelli, N. J. (2004). The mid-domain effect and species richness patterns: what have we learned so far?. The American Naturalist, 163(3), E1-E23.
32. Colwell, R. K., Brehm, G., Cardelús, C. L., Gilman, A. C., & Longino, J. T. (2008). Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. Science, 322(5899), 258-261.
33. Crawford, A. J. (2003). Relative rates of nucleotide substitution in frogs. Journal of Molecular Evolution, 57(6), 636-641.
34. Cretu Stancu, M., Van Roosmalen, M. J., Renkens, I., Nieboer, M. M., Middelkamp, S., De Ligt, J., ... & Kloosterman, W. P. (2017). Mapping and phasing of structural variation in patient genomes using nanopore sequencing. Nature communications, 8(1), 1326
35. Crothers JH. (1998). A hot summer, cold winters, and the geographical limit of Trochocochlea lineata in Somerset. Hydrobiologia 378:133–41

36. Currie, D. J. (1991). Energy and large-scale patterns of animal-and plant-species richness. The American Naturalist, 137(1), 27-49.
37. Daco, L., Matthies, D., Hermant, S., & Colling, G. (2022). Genetic diversity and differentiation of populations of Anthyllis vulneraria along elevational and latitudinal gradients. Ecology and Evolution, 12(8), e9167
38. Darst, C. R., & Cannatella, D. C. (2004). Novel relationships among hyloid frogs inferred from 12S and 16S mitochondrial DNA sequences. Molecular phylogenetics and evolution, 31(2), 462-475.
39. Davies, T. J. (2006). Evolutionary ecology: when relatives cannot live together. Current Biology, 16(16), R645-R647.
40. da Silva, F. K. G., de Faria Lopes, S., Lopez, L. C. S., de Melo, J. I. M., & Trovão, D. M. D. B. M. (2014). Patterns of species richness and conservation in the Caatinga along elevational gradients in a semiarid ecosystem. Journal of Arid Environments, 110, 47-52.
41. Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics, 12(7), 499-510.
42. Deagle, B. E., Thomas, A. C., McInnes, J. C., Clarke, L. J., Vesterinen, E. J., Clare, E. L., ... & Eveson, J. P. (2019). Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data?. Molecular ecology, 28(2), 391-406
43. Desjardins, P., & Conklin, D. (2010). NanoDrop microvolume quantitation of nucleic acids. Journal of visualized experiments: JoVE, (45), 2565.
44. Dhib, J., Comas, M., Moreno-Rueda, G., & Selmi, S. (2024). Retreat site occupation and short-term movements in an oasis-dwelling frog in relation to temperature-humidity conditions. Journal of Arid Environments, 222, 105158.
45. Dinesh, K. P., Radhakrishnan, C., Channakeshavamurthy, B. H., & Kulkarni, N. U. (2015). A Checklist of Amphibians of India, (January), 1–13
46. Dobzhansky, T. (1950). Evolution in the tropics. American scientist, 38(2), 209-221.
47. Dolmen, D., & Seland, J. (2016). How fast do amphibians disperse? Introductions, distribution and dispersal of the common frog Rana temporaria and the common toad Bufo bufo on a coastal island in Central Norway. Fauna norvegica, 36, 33-46.
48. Edwards MS, Hernandez-Carmona G. (2005). Delayed recovery of giant kelp near its southern range limit in the North Pacific following El Niño. Mar. Biol. 147:273–79

49. Thiel-Egenter, C., Gugerli, F., Alvarez, N., Brodbeck, S., Cieślak, E., Colli, L., ... & IntraBioDiv Consortium. (2009). Effects of species traits on the genetic diversity of high-mountain plants: a multi-species study across the Alps and the Carpathians. Global Ecology and Biogeography, 18(1), 78-87.
50. Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity, 107(1), 1-15.
51. Elsen, P. R., Tingley, M. W., Kalyanaraman, R., Ramesh, K., & Wilcove, D. S. (2017). The role of competition, ecotones, and temperature in the elevational distribution of Himalayan birds. Ecology, 98(2), 337-348
52. Espada, R., Zarevski, N., Dramé-Maigné, A., & Rondelez, Y. (2022). Accurate gene consensus at low nanopore coverage. GigaScience, 11, giac102
53. Faivovich, J., Haddad, C. F., Garcia, P. C., Frost, D. R., Campbell, J. A., & Wheeler, W. C. (2005). Systematic review of the frog family Hylidae, with special reference to Hylinae: phylogenetic analysis and taxonomic revision. Bulletin of the American Museum of natural History, 2005(294), 1-240.
54. Feijo, A., Wen, Z., Cheng, J., Ge, D., Xia, L., & Yang, Q. (2019). Divergent selection along elevational gradients promotes genetic and phenotypic disparities among small mammal populations. Ecology and Evolution, 9(12), 7080-7095.
55. Felkel, S., Tremetsberger, K., Moser, D., Dohm, J. C., Himmelbauer, H., & Winkler, M. (2023). Genome-environment associations along elevation gradients in two snowbed species of the North-Eastern Calcareous Alps. BMC Plant Biology, 23(1), 203.
56. Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. Nature Reviews Microbiology, 15(10), 579–590
57. Freeman, B. G., & Montgomery, G. (2016). Interspecific aggression by the Swainson's Thrush (Catharus ustulatus) may limit the distribution of the threatened Bicknell's Thrush (Catharus bicknelli) in the Adirondack Mountains. The Condor: Ornithological Applications, 118(1), 169-178.
58. Freeman, B. G., Strimas-Mackey, M., & Miller, E. T. (2022). Interspecific competition limits bird species' ranges in tropical mountains. Science, 377(6604), 416-420
59. Fritz, S. A., Eronen, J. T., Schnitzler, J., Hof, C., Janis, C. M., Mulch, A., ... & Graham, C. H. (2016). Twenty-million-year relationship between mammalian diversity and primary productivity. Proceedings of the National Academy of Sciences, 113(39), 10908-10913.

60. Funk, W. C., Greene, A. E., Corn, P. S., & Allendorf, F. W. (2005). High dispersal in a frog species suggests that it is vulnerable to habitat fragmentation. Biology letters, 1(1), 13-16.
61. Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular ecology, 18(22), 4541-4550.
62. Gaston, K. J. (1991). How Large Is a Species' Geographic Range? Oikos, 61(3), 434 – 438.
63. Gaston, K. J. (2000). Global patterns in biodiversity. Nature, 405(6783), 220-227.
64. Gass, J., Miller, A. J., Sheets, C., Long, M., & Voyles, J. (2024). High relative humidity and temperature limit disease development and mortality in golden frogs of Panama, Atelopus zeteki, infected with Batrachochytrium dendrobatidis. Evolutionary Ecology, 38(1), 141-156.
65. Gebrehiwot, K., Demissew, S., Woldu, Z., Fekadu, M., Desalegn, T., & Teferi, E. (2019). Elevational changes in vascular plants richness, diversity, and distribution pattern in Abune Yosef mountain range, Northern Ethiopia. Plant diversity, 41(4), 220-228.
66. Germain, R. M., Weir, J. T., & Gilbert, B. (2016). Species coexistence: macroevolutionary relationships and the contingency of historical interactions. Proceedings of the Royal Society B: Biological Sciences, 283(1827), 20160047
67. Giehl, E.L.H. & Jarenkow, J.A. (2012) Niche conservatism and the differences in species richness at the transition of tropical and subtropical climates in South America. Ecography, 35, 933–943
68. Giordano, A. R., Ridenhour, B. J., & Storfer, A. (2007). The influence of altitude and topography on genetic structure in the long-toed salamander (Ambystoma macrodactulym). Molecular Ecology, 16(8), 1625-1637
69. Gizachew, G. T. (2021). Spatial-temporal and factors influencing the distribution of biodiversity: A Review. Scientific Reports in Life Sciences, 2(4), 1-19.
70. Godoy, O., Kraft, N. J., & Levine, J. M. (2014). Phylogenetic relatedness and the determinants of competitive outcomes. Ecology letters, 17(7), 836-844
71. Goebel, A. M., Donnelly, J. M., & Atz, M. E. (1999). PCR primers and amplification methods for 12S ribosomal DNA, the control region, cytochrome oxidase I, and cytochromebin bufonids and other frogs, and an overview of PCR primers which have amplified DNA in amphibians successfully. Molecular Phylogenetics and Evolution, 11(1), 163-199.

72. Golan, D., Erlich, Y., & Rosset, S. (2012). Weighted pooling—practical and cost-effective techniques for pooled high-throughput sequencing. Bioinformatics, 28(12), i197-i206
73. Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. C., & McCombie, W. R. (2015). Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome research, 25(11), 1750-1756
74. GRAYBEAL, A. (1997). Phylogenetic relationships of bufonid frogs and tests of alternate macroevolutionary hypotheses characterizing their radiation. Zoological journal of the Linnean Society, 119(3), 297-338.
75. Haendiges, J., Timme, R., Ramachandran, P., & Balkey, M. (2020). DNA Quantification Using the Qubit Fluorometer.
76. Guo, Q., Kelt, D. A., Sun, Z., Liu, H., Hu, L., Ren, H., & Wen, J. (2013). Global variation in elevational diversity patterns. Sci Rep 3: 3007.
77. Hahn, T., Kettle, C. J., Ghazoul, J., Frei, E. R., Matter, P., & Pluess, A. R. (2012). Patterns of genetic variation across altitude in three plant species of semi-dry grasslands
78. Hargreaves, A.L., Samis, K.E. & Eckert, C.G. (2014) Are species' range limits simply niche limits writ large? A review of transplant experiments beyond the range. The American Naturalist, 183, 157–173
79. Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guégan, J. F., Kaufman, D. M., ... & Turner, J. R. (2003). Energy, water, and broad-scale geographic patterns of species richness. Ecology, 84(12), 3105-3117.
80. Heaney, L. R. (2001). Small mammal diversity along elevational gradients in the Philippines: an assessment of patterns and hypotheses. Global ecology and Biogeography, 10(1), 15-39.
81. Hebert, P. D., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1512), 313-321.
82. Hebert, P. D., & Gregory, T. R. (2005). The promise of DNA barcoding for taxonomy. Systematic biology, 54(5), 852-859.
83. Hedges, S. B. (1994). Molecular evidence for the origin of birds. Proceedings of the National Academy of Sciences, 91(7), 2621-2624.

84. Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., ... & Rokhsar, D. S. (2010). The genome of the Western clawed frog Xenopus tropicalis. Science, 328(5978), 633-636.
85. Hersteinsson P, Macdonald DW. (1992). Interspecific competition and the geographical distribution of red and arctic foxes Vulpes vulpes and Alopex lagopus. Oikos 64:505–15
86. Hoegg, S., Vences, M., Brinkmann, H., & Meyer, A. (2004). Phylogeny and comparative substitution rates of frogs inferred from sequences of three nuclear genes. Molecular Biology and Evolution, 21(7), 1188-1200.
87. Holt, R.D. and Keitt, T.H. (2005), Species' borders: a unifying theme in ecology. Oikos, 108: 3 – 6.
88. Holt, B. G., Lessard, J. P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., ... & Rahbek, C. (2013). An update of Wallace's zoogeographic regions of the world. Science, 339(6115), 74-78.
89. Hwang, U. W., & Kim, W. (1999). General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics. The Korean journal of parasitology, 37(4), 215.
90. Jain, M., Tyson, J. R., Loose, M., Ip, C. L., Eccles, D. A., O'Grady, J., ... & Reference Consortium. (2017). MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9. 0 chemistry. F1000Research, 6, 760
91. Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., ... & Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. Nature biotechnology, 36(4), 338-345
92. Jackson, S. T., & Overpeck, J. T. (2000). Responses of plant populations and communities to environmental changes of the late Quaternary. Paleobiology, 26(S4), 194-220.
93. Jankowski, J. E., Robinson, S. K., & Levey, D. J. (2010). Squeezed at the top: Interspecific aggression may constrain elevational ranges in tropical birds. Ecology, 91(7), 1877-1884
94. Janzen, D. H. (1967). Why mountain passes are higher in the tropics. The American Naturalist, 101(919), 233-249.
95. Jetz, W, Thomas, GH, Joy, JB, Hartmann, K and Mooers, AO (2012) The global diversity of birds in space and time. Nature 491, 444–448.

96. Kerkhoff, A. J., Moriarty, P. E., & Weiser, M. D. (2014). The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. Proceedings of the National Academy of Sciences, 111(22), 8125-8130.
97. Kirkpatrick M, Barton NH. (1997). Evolution of a species' range. Am. Nat. 150:1–23
98. Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences, 86(16), 6196-6200.
99. Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome research, 27(5), 722-736
100. Körner, C. (2007). The use of 'altitude'in ecological research. Trends in ecology & evolution, 22(11), 569-574.
101. Kozak, K.H. & Wiens, J.J. (2006) Does niche conservatism drive speciation? A case study in North American salamanders. Evolution, 60, 2604–2621
102. Kozak, K.H. & Wiens, J.J. (2010) Niche conservatism drives elevational diversity patterns in Appalachian salamanders. The American Naturalist, 176, 40–54
103. Kraft, N. J., Adler, P. B., Godoy, O., James, E. C., Fuller, S., & Levine, J. M. (2015). Community assembly, coexistence and the environmental filtering metaphor. Functional ecology, 29(5), 592-599.
104. Krosby, M., Wilsey, C. B., McGuire, J. L., Duggan, J. M., Nogeire, T. M., Heinrichs, J. A., ... & Lawler, J. J. (2015). Climate-induced range overlap among closely related species. Nature Climate Change, 5(9), 883-886
105. Laver, T., Harrison, J., O'neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., & Studholme, D. J. (2015). Assessing the performance of the oxford nanopore technologies minion. Biomolecular detection and quantification, 3, 1-8
106. Lau, Q., Igawa, T., Ogino, H., Katsura, Y., Ikemura, T., & Satta, Y. (2020). Heterogeneity of synonymous substitution rates in the Xenopus frog genome. PLoS One, 15(8), e0236515.
107. Lee-Yaw, J.A., Kharouba, H.M., Bontrager, M., Mahony, C., Csergő, A.M., Noreen, A.M.E. et al. (2016) A synthesis of transplant experiments and ecological niche models suggests that range limits are often niche limits. Ecology Letters, 19, 710–722

108. Lenoir, J., Gégout, J. C., Marquet, P. A., de Ruffray, P., & Brisse, H. (2008). A significant upward shift in plant species optimum elevation during the 20th century. science, 320(5884), 1768-1771.

109. Li, H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics, 32(14), 2103-2110

110. Liao, J., Chen, S., Liu, P., Fontaneto, D., & Han, B. P. (2024). Environmental selection and gene flow jointly determine the population genetic diversity and structure of Diaphanosoma dubium along a watershed elevation. Global Ecology and Conservation, 49, e02773.

111. Lillywhite, H. B., & Navas, C. A. (2006). Animals, energy, and water in extreme environments: perspectives from Ithala 2004. Physiological and Biochemical Zoology, 79(2), 265-273.

112. Loman, N. J., Quick, J., & Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. Nature methods, 12(8), 733-735

113. Lomolino, M. V. (2001). Elevation gradients of species-density: historical and prospective views. Global Ecology and biogeography, 10(1), 3-13.

114. Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. Ecology letters, 11(10), 995-1003.

115. Low, V. L., Adler, P. H., Takaoka, H., Ya'cob, Z., Lim, P. E., Tan, T. K., ... & Sofian-Azirun, M. (2014). Mitochondrial DNA markers reveal high genetic diversity but low genetic differentiation in the black fly Simulium tani Takaoka & Davies along an elevational gradient in Malaysia. PLoS One, 9(6), e100512

116. Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. Genomics, proteomics & bioinformatics, 14(5), 265-279.

117. Lundblad, C. G., & Conway, C. J. (2020). Variation in selective regimes drives intraspecific variation in life-history traits and migratory behaviour along an elevational gradient. Journal of Animal Ecology, 89(2), 397-411

118. Luquet, É., Léna, J. P., Miaud, C., & Plénet, S. (2015). Phenotypic divergence of the common toad (Bufo bufo) along an altitudinal gradient: evidence for local adaptation. Heredity, 114(1), 69-79.

119. MacArthur, R. H. (1984). Geographical ecology: patterns in the distribution of species. Princeton University Press.

120. Maitra, A., Pandit, R., Mungee, M., & Athreya, R. (2022). Testing a theoretical framework for the environment-species abundance paradigm: a new approach to the abundant centre hypothesis. bioRxiv, 2022-01.
121. Manish, K. (2021). Species richness, phylogenetic diversity and phylogenetic structure patterns of exotic and native plants along an elevational gradient in the Himalaya. Ecological Processes, 10(1), 64
122. Marathe, A., Priyadarsanan, D. R., Krishnaswamy, J., & Shanker, K. (2020). Spatial and climatic variables independently drive elevational gradients in ant species richness in the Eastern Himalaya. PLoS One, 15(1), e0227628.
123. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., ... & Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature, 437(7057), 376-380.
124. Marin, J., Rapacciuolo, G., Costa, G. C., Graham, C. H., Brooks, T. M., Young, B. E., ... & Hedges, S. B. (2018). Evolutionary time drives global tetrapod diversity. Proceedings of the Royal Society B: Biological Sciences, 285(1872).
125. Massad, T. J., Abrão, O. J., António, H., Chechene, A., Soares C. Tenente, B., André, A., ... & Naskrecki, P. (2024). Ecosystem-wide responses to fire and large mammal herbivores in an African savanna. Biotropica, 56(4), e13338
126. Mayfield, M. M., & Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. Ecology letters, 13(9), 1085-1093
127. McCain, C. M. (2009). Global analysis of bird elevational diversity. Global Ecology and Biogeography, 18(3), 346-360.
128. McCain, C. M., & Grytnes, J. A. (2010). Elevational gradients in species richness. Encyclopedia of life sciences, 15, 1-10.
129. McGill, B. J. (2010). Towards a unification of unified theories of biodiversity. Ecology letters, 13(5), 627-642.
130. Meng, H., Wei, X., & Jiang, M. (2019). Contrasting elevational patterns of genetic variation in Euptelea pleiospermum along mountains at the core and edges of its latitudinal range. Plant Ecology, 220(1), 13-28.
131. Metzker, M. L. (2010). Sequencing technologies—the next generation. Nature reviews genetics, 11(1), 31-46.

132. Miraldo, A., Li, S., Borregaard, M. K., Flórez-Rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., ... & Nogués-Bravo, D. (2016). An Anthropocene map of genetic diversity. Science, 353(6307), 1532-1535

133. Mishra, A. (2015). Genetic structure in genus Philautus (Gistel, 1848) Family: Rhacophoridae, Order: Anura) along an altitudinal gradient in Eastern Himalayas, (March), 1–53.

134. Mittelbach, G. G., Schemske, D. W., Cornell, H. V., Allen, A. P., Brown, J. M., Bush, M. B., ... & Turelli, M. (2007). Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecology letters, 10(4), 315-331

135. Moran, E. V., Reid, A., & Levine, J. M. (2017). Population genetics and adaptation to climate along elevation gradients in invasive Solidago canadensis. PLoS One, 12(9), e0185539.

136. Mullis, K. B., & Faloona, F. A. (1987). [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In Methods in enzymology (Vol. 155, pp. 335-350). Academic Press.

137. Mungee, M., & Athreya, R. (2021). Intraspecific trait variability and community assembly in hawkmoths (Lepidoptera: Sphingidae) across an elevational gradient in the eastern Himalayas, India. Ecology and Evolution, 11(6), 2471-2487.

138. Mungee, M., Pandit, R., & Athreya, R. (2021). Taxonomic scale dependency of Bergmann's patterns: a cross-scale comparison of hawkmoths and birds along a tropical elevational gradient. Journal of Tropical Ecology, 37(6), 302-312.

139. Mungee, M., Pandit, R., & Athreya, R. (2025). Strong seasonal and elevational dynamics in community assembly mechanisms of tropical montane birds in the Eastern Himalayas. Tropical Ecology, 66(4), 559-572.

140. Muster, C. (2002). Substitution patterns in congeneric arachnid species in the northern Alps. Diversity and Distributions, 8(2), 107-121

141. Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature, 403(6772), 853-858.

142. Nardin, M., Musch, B., Rousselle, Y., Guérin, V., Sanchez, L., Rossi, J. P., ... & Rozenberg, P. (2015). Genetic differentiation of European larch along an altitudinal gradient in the French Alps. Annals of Forest Science, 72(5), 517-527

143. Navalón, G., Bright, J. A., Marugán-Lobón, J., & Rayfield, E. J. (2019). The evolutionary relationship among beak shape, mechanical advantage, and feeding ecology in modern birds. Evolution, 73(3), 422-435

144. Ohsawa, T., & Ide, Y. (2008). Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. Global Ecology and Biogeography, 17(2), 152-163.

145. Olson, D. M., & Dinerstein, E. (1998). The Global 200: a representation approach to conserving the Earth's most biologically valuable ecoregions. Conservation biology, 12(3), 502-515.

146. Orme, C. D. L., Davies, R. G., Burgess, M., Eigenbrod, F., Pickup, N., Olson, V. A., ... & Owens, I. P. (2005). Global hotspots of species richness are not congruent with endemism or threat. Nature, 436(7053), 1016-1019.

147. Palumbi, S.R., A. Martin, W.O. McMillan, L. Stice, and G. Grabowski. 1991. The simple fool's guide to PCR, version 2.0. Privately published

148. Pandey, N., Khanal, L., & Chalise, M. K. (2020). Correlates of avifaunal diversity along the elevational gradient of Mardi Himal in Annapurna Conservation Area, Central Nepal. Avian Research, 11(1), 31

149. Paudel, P. K., Sipos, J., & Brodie, J. F. (2018). Threatened species richness along a Himalayan elevational gradient: quantifying the influences of human population density, range size, and geometric constraints. BMC ecology, 18(1), 6.

150. Parmesan C. (2006). Ecological and evolutionary responses to recent climate change. Annu. Rev. Ecol. Evol. Syst. 37:637–69

151. Paul, A., Khan, M.L., Arunachalam, A. and Arunachalam, K., 2005. Biodiversity and conservation of rhododendrons in Arunachal Pradesh in the Indo-Burma biodiversity hotspot. Current Science, 89(4), pp.623-634

152. Payne, A., Holmes, N., Rakyan, V., & Loose, M. (2019). BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. Bioinformatics, 35(13), 2193-2198

153. Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. Quarterly Review of Biology, 78, 419–433

154. Pfennig, K., & Pfennig, D. (2009). Character displacement: ecological and reproductive responses to a common evolutionary problem. The quarterly review of biology, 84(3), 253-276.

155. Piaggio, A. J., Navo, K. W., & Stihler, C. W. (2009). Intraspecific comparison of population structure, genetic diversity, and dispersal among three subspecies of Townsend's big-eared bats, Corynorhinus townsendii townsendii, C. t. pallescens, and the endangered C. t. virginianus. Conservation Genetics, 10(1), 143-159

156. Pianka, E. R. (1966). Latitudinal gradients in species diversity: a review of concepts. The American Naturalist, 100(910), 33-46.

157. Picelli, S., Björklund, Å. K., Reinius, B., Sagasser, S., Winberg, G., & Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome research, 24(12), 2033-2040.

158. Pigot, A. L., Jetz, W., Sheard, C., & Tobias, J. A. (2018). The macroecological dynamics of species coexistence in birds. Nature Ecology & Evolution, 2(7), 1112-1119.

159. Polato, N. R., Gray, M. M., Gill, B. A., Becker, C. G., Casner, K. L., Flecker, A. S., ... & Zamudio, K. R. (2017). Genetic diversity and gene flow decline with elevation in montane mayflies. Heredity, 119(2), 107-116

160. Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T., Alström, P., ... & Mohan, D. (2014). Niche filling slows the diversification of Himalayan songbirds. Nature, 509(7499), 222-225.

161. Rahbek, C. (1995). The elevational gradient of species richness: a uniform pattern?. Ecography, 200-205.

162. Rang, F.J., Kloosterman, W.P. & de Ridder, J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biol 19, 90 (2018)

163. Rasmussen, P. C., & Anderton, J. C. (2005). Birds of south Asia: the Ripley guide (Vol. 2, pp. 1-378).

164. Reisch, C., & Rosbakh, S. (2021). Patterns of genetic variation in European plant species depend on altitude. Diversity and Distributions, 27(1), 157-163.

165. Ricklefs, R. E. (2006). Evolutionary diversification and the origin of the diversity–environment relationship. Ecology, 87(sp7), S3-S13.

166. Robin, V. V., Sinha, A., & Ramakrishnan, U. (2010). Ancient geographical gaps and paleo-climate shape the phylogeography of an endemic bird in the sky islands of southern India. PLoS One, 5(10), e13321

167. Rohde, K. (1992). Latitudinal gradients in species diversity: the search for the primary cause. Oikos, 514-527.

168. Rosenzweig, M. L. (1996). Species diversity in space and time (Vol. 344). Cambridge: Cambridge university press.
169. Rosenzweig, M. L., & Sandlin, E. A. (1997). Species diversity and latitudes: listening to area's signal. Oikos, 172-176.
170. Sambrook, J., & Russell, D. W. (2001). Molecular Cloning: Ch. 8. In Vitro amplification of DNA by the polymerase chain reaction (Vol. 2). Cold Spring Harbor Laboratory Press.
171. Sanders, N. J., Moss, J., & Wagner, D. (2003). Patterns of ant species richness along elevational gradients in an arid ecosystem. Global ecology and biogeography, 12(2), 93-102.
172. Sanford, E., Holzman, S.B., Haney, R.A., Rand, D.M., Bertness, M.D., (2006). Larval tolerance, gene flow, and the northern geographic range limit of fiddler crabs. Ecology. 87, 2882–2894
173. Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of molecular biology, 94(3), 441-448
174. Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the national academy of sciences, 74(12), 5463-5467.
175. Schemske, D. W. (2009). CHAPTER TWELVE Biotic interactions and speciation in the tropics. Speciation and patterns of diversity.
176. Schumm, M., White, A. E., Supriya, K., & Price, T. D. (2020). Ecological limits as the driver of bird species richness patterns along the east Himalayan elevational gradient. The American Naturalist, 195(5), 802-817
177. Schoener, T. W. (2010). The MacArthur-Wilson Equilibrium Model. The theory of island biogeography revisited, 52-87.
178. Semlitsch, R. D. (2008). Differentiating migration and dispersal processes for pond-breeding amphibians. The Journal of wildlife management, 72(1), 260-267.
179. Sharma, N., Singh, D., Sharma, S., & Ansari, A. (2023). A preliminary assessment of Odonata (dragonflies & damselflies) across an elevation gradient–insights from Shiwaliks to Alpines, northwestern Himalaya, India. Journal of Threatened Taxa, 15(7), 23545-23556.
180. Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. Nature biotechnology, 26(10), 1135-1145.
181. Shi, M. M., Michalski, S. G., Chen, X. Y., & Durka, W. (2011). Isolation by elevation: genetic structure at neutral and putatively non-neutral loci in a dominant tree of subtropical forests, Castanopsis eyrei. PLoS One, 6(6), e21302.

182. Shimada, T., Matsui, M., Yambun, P., & Sudin, A. (2011). A taxonomic study of Whitehead's torrent frog, Meristogenys whiteheadi, with descriptions of two new species (Amphibia: Ranidae). Zoological Journal of the Linnean Society, 161(1), 157-183.

183. Shokralla, S., Gibson, J. F., Nikbakht, H., Janzen, D. H., Hallwachs, W., & Hajibabaei, M. (2014). Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Molecular ecology resources, 14(5), 892-901

184. Simbolo, M., Gottardi, M., Corbo, V., Fassan, M., Mafficini, A., Malpeli, G., ... & Scarpa, A. (2013). DNA qualification workflow for next generation sequencing of histopathological samples. PloS one, 8(6), e62692.

185. Snider, M. H., Helgen, K. M., Young, H. S., Agwanda, B., Schuttler, S., Titcomb, G. C., ... & Kays, R. (2024). Shifting mammal communities and declining species richness along an elevational gradient on Mount Kenya. Ecology and Evolution, 14(4), e11151

186. Spranger, R. R., Raffel, T. R., & Sinervo, B. R. (2024). Canopy coverage, light, and moisture affect thermoregulatory trade-offs in an amphibian breeding habitat. Journal of Thermal Biology, 122, 103864.

187. Srinivasan, U., Raza, R. H., & Quader, S. (2010). The nuclear question: rethinking species importance in multi-species animal groups. Journal of animal ecology, 79(5), 948-954.

188. Srinivasan, U., Elsen, P. R., Tingley, M. W., & Wilcove, D. S. (2018). Temperature and competition interact to structure Himalayan bird communities. Proceedings of the Royal Society B: Biological Sciences, 285(1874), 20172593

189. Steinbauer, M. J., Field, R., Grytnes, J. A., Trigas, P., Ah-Peng, C., Attorre, F., ... & Beierkuhnlein, C. (2016). Topography-driven isolation, speciation and a global increase of endemism with elevation. Global Ecology and Biogeography, 25(9), 1097-1107.

190. Stephens, P.R. & Wiens, J.J. (2009) Bridging the gap between community ecology and historical biogeography: niche conservatism and community structure in emydid turtles. Molecular Ecology, 18, 4664–4679

191. Stevens, G. C. (1992). The elevational gradient in altitudinal range: an extension of Rapoport's latitudinal rule to altitude. The American Naturalist, 140(6), 893-911.

192. Stock, W., Rousseau, C., Dierickx, G., D'hondt, S., Amadei Martínez, L., Dittami, S. M., ... & De Clerck, O. (2024). Breaking free from references: a consensus-based approach for community profiling with long amplicon nanopore data. Briefings in Bioinformatics, 26(1)

193. Sun, Y. B., Xiong, Z. J., Xiang, X. Y., Liu, S. P., Zhou, W. W., Tu, X. L., ... & Zhang, Y. P. (2015). Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes. Proceedings of the National Academy of Sciences, 112(11), E1257-E1262.

194. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular ecology, 21(8), 2045-2050

195. TANG, Z. Y., & FANG, J. Y. (2004). A review on the elevational patterns of plant species diversity. Biodiversity Science, 12(1), 20.

196. Terborgh, J. (1977). Bird species diversity on an Andean elevational gradient. Ecology, 58(5), 1007-1019.

197. Tilman, D. (1982). Resource competition and community structure (No. 17). Princeton university press.

198. Titus, T. A., & Larson, A. (1996). Molecular phylogenetics of desmognathine salamanders (Caudata: Plethodontidae): a reevaluation of evolution in ecology, life history, and morphology. Systematic Biology, 45(4), 451-472.

199. Tyson, J. R., O'Neil, N. J., Jain, M., Olsen, H. E., Hieter, P., & Snutch, T. P. (2018). MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome. Genome research, 28(2), 266-274

200. Umaña, M. N., & Swenson, N. G. (2019). Intraspecific variation in traits and tree growth along an elevational gradient in a subtropical forest. Oecologia, 191(1), 153-164.

201. Valeur, B., & Berberan-Santos, M. N. (2013). Molecular fluorescence: principles and applications. John Wiley & Sons.

202. van der Meijden, A., Vences, M., Hoegg, S., Boistel, R., Channing, A., & Meyer, A. (2007). Nuclear gene phylogeny of narrow-mouthed toads (Family: Microhylidae) and a discussion of competing hypotheses concerning their biogeographical origins. Molecular Phylogenetics and Evolution, 44(3), 1017-1030.

203. van Els, P., Herrera-Alsina, L., Pigot, A. L., & Etienne, R. S. (2021). Evolutionary dynamics of the elevational diversity gradient in passerine birds. Nature Ecology & Evolution, 5(9), 1259-1265.

204. Vázquez G., & Givnish, T. J. (1998). Altitudinal gradients in tropical forest composition, structure, and diversity in the Sierra de Manantlán. Journal of ecology, 999-1020.

205. Violle, C., Nemergut, D. R., Pu, Z., & Jiang, L. (2011). Phylogenetic limiting similarity and competitive exclusion. Ecology letters, 14(8), 782-787

206. Wallace, A. R. (1876, January). Lecture on the comparative antiquity of continents, as indicated by the distribution of living and extinct animals. In Proceedings of the Royal Geographical Society of London (Vol. 21, No. 6, pp. 505-535). Royal Geographical Society (with the Institute of British Geographers), Wiley.

207. Whittaker, R. H. (1967). Gradient analysis of vegetation.

208. Whittaker, R. H., & Niering, W. A. (1975). Vegetation of the Santa Catalina Mountains, Arizona. V. Biomass, production, and diversity along the elevation gradient. Ecology, 56(4), 771-790.

209. Wick RR, Judd LM, Holt KE. Comparison of Oxford nanopore basecalling tools. Zenodo 2018. https://zenodo.org/record/1188469#.Ww0upI-cGM8

210. Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. Trends in ecology & evolution, 19(12), 639-644.

211. Wiens, J. J., Fetzner Jr, J. W., Parkinson, C. L., & Reeder, T. W. (2005). Hylid frog phylogeny and sampling strategies for speciose clades. Systematic biology, 54(5), 778-807.

212. Wiens, J.J., Graham, C.H., Moen, D.S., Smith, S.A. & Reeder, T.W. (2006) Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. The American Naturalist, 168, 579–596

213. Wilkinson, J. A., Matsui, M., & Terachi, T. (1996). Geographic variation in a Japanese tree frog (Rhacophorus arboreus) revealed by PCR-aided restriction site analysis of mtDNA. Journal of Herpetology, 30(3), 418-423.

214. Zhou, Y., Ochola, A. C., Njogu, A. W., Boru, B. H., Mwachala, G., Hu, G., ... & Wang, Q. (2019). The species richness pattern of vascular plants along a tropical elevational gradient and the test of elevational Rapoport's rule depend on different life-forms and phytogeographic affinities. Ecology and Evolution, 9(8), 4495-4503.

215. Zomer, R., Ustin, S., & Ives, J. (2002). Using satellite remote sensing for DEM extraction in complex mountainous terrain: landscape analysis of the Makalu Barun National Park of eastern Nepal. International Journal of Remote Sensing, 23(1), 125-143.

216. Zormpa G, O., Wilhelmi, S., Vucetic, B., Ciocîrlan, M. I. C., Mueller, M., Ciocîrlan, E., ... & Budde, K. B. (2025). Genetic diversity and fine-scale spatial genetic structure of European beech populations along an elevational gradient. Heredity, 134(8), 451-460.

217. Zou, J. Y., Luo, Y. H., Burgess, K. S., Tan, S. L., Zheng, W., Fu, C. N., ... & Gao, L. M. (2021). Joint effect of phylogenetic relatedness and trait selection on the elevational distribution of Rhododendron species. Journal of Systematics and Evolution, 59(6), 1244-1255.