Template/Pro forma for Submission

NMHS-FINAL TECHNICAL REPORT (FTR)

Demand-Driven Action Research Project Grant

NMHS Reference No.:	NMHS/SG-2016/011

Date of Submission:	1	4	1	1	2	0	1	9
	d	**d**	**m**	**m**	**y**	**y**	**y**	**y**

PROJECT TITLE (IN CAPITAL)

CONSERVATION STRATEGIES FOR *Taxus wallichiana* AND *Ulmus wallichiana* BY DNA MARKERS AND GEOSPATIAL TECHNOLOGIES

Project Duration: *from* (**01.04.2016**) *to* (**31.03.2019**).

***Submitted to*:**
Er. Kireet Kumar
Scientist 'G' and Nodal Officer, NMHS-PMU
National Mission on Himalayan Studies, GBPNIHESD HQs
Ministry of Environment, Forest & Climate Change (MoEF&CC), New Delhi
E-mail: nmhspmu2016@gmail.com; kireet@gbpihed.nic.in; subratabose@nic.in

***Submitted by*:**
[*Dr. Pankaj Bhardwaj*]
[*Dept. of Plant sciences, Central University of Punjab, Bathinda. Punjab*]
[*Contact No.: 9501686709*]
[E-mail: pankajihbt@gmail.com.]

GENERAL INSTRUCTIONS:

1. The Final Technical Report (FTR) has to commence from the date of start of the Project (as per the Sanction Order issued at the start of the project) till its completion. Each detail has to comply with the NMHS Sanction Order.
2. The FTR should be neatly typed (in Arial with font size 11 with 1.5 spacing between the lines) with all details as per the enclosed format for direct reproduction by photo-offset process. Colored Photographs (4-5 good action photographs), tables and graphs should be accommodated within the report or should be annexed with captions. Sketches and diagrammatic illustrations may also be given giving step-by-step details about the methodology followed in technology development/modulation, transfer and training. Any correction or rewriting should be avoided. Please give information under each head in serial order.
3. Training/ Capacity Building Manuals (with details contents of training programme technical details and techniques involved) or any such display material related to project activities along with slides, charts, photographs should be brought at the venue of the Annual Monitoring & Evaluation (M&E) Workshop and sent at the NMHS-PMU, GBPNIHESD HQs, Kosi-Katarmal, Almora 263643, Uttarakhand. In all Knowledge Products, the Grant/ Fund support of the NMHS should be duly acknowledged.
4. The FTR Format is in sync with many other essential requirements and norms desired by the Govt. of India time to time, so each section of the NMHS-FTR needs to duly filled by the proponent and verified by the Head of the Lead Implementing Organization/ Institution/ University.
5. Five (5) bound hard copies of the Project Final Technical Report (FTR) and a soft copy should be submitted to the **Nodal Officer, NMHS-PMU, GBPNIHESD HQs, Kosi-Katarmal, Almora, Uttarakhand**.

The FTR is to be submitted into following two parts:

Part A – Project Summary Report

Part B – Project Detailed Report

Following Financial and other necessary documents/certificates need to be submitted along with Final Technical Report (FTR):

Annexure I	**Consolidated and Audited Utilization Certificate (UC) & Statement of Expenditure (SE)**, including interest earned for the last Fiscal year including the duly filled GFR-19A (with year-wise break-up)
Annexure II	**Consolidated Interest Earned Certificate**
Annexure III	**Consolidated Assets Certificate** showing the cost of the equipment in Foreign and Indian currency, Date of Purchase, etc. (with break-up as per the NMHS Sanction Order and year wise).
Annexure IV	**List of all the equipment, assets and peripherals** purchased through the NMHS grant with current status of use including location of deployment.
Annexure V	**Consolidated Manpower Certificate** and **Direct Benefit Transfer (DBT) Details** showing the education background, i.e. NET/GATE etc. qualified or not, Date of joining and leaving, Salary paid per month and per annum (with break up as per the Sanction Order and year-wise).
Annexure VII	**Refund of any unspent balance as Demand Draft in favor of DDO, GBPNIHESD** payable at GBPNIHESD, Kosi-Katarmal, Almora, Uttarakhand.
Annexure VIII	**Details of Technology Transfer and Intellectual Property Rights developed.**

DSL: Date of Sanction Letter

3	1	0	3	**2**	**0**	**1**	**6**
d	d	m	m	y	y	y	y

DPC: Date of Project completion

3	1	0	9	**2**	**0**	**1**	**9**
d	d	m	m	y	y	y	y

Part A: Project Summary Report

1. Project Description

i.	Project Reference No.								
ii.	Type of Project	Small Grant	✓	Medium Grant		Large Grant			
iii.	Project Title	Conservation strategies for *Taxus wallichiana* and *Ulmus wallichiana* by DNA markers and geospatial technologies							
iv.	State under which Project is Sanctioned	**Punjab**							
v.	Project Sites (IHR States covered) (Maps to be attached)	J&K, HP and UK							
vi.	Scale of Project Operation	Local		Regional		Pan-Himalayan	✓		
vii.	Total Budget/ Outlay of the Project	3494600/- (in Cr)							
viii.	Lead Agency	Central University of Punjab							
	Principal Investigator (PI)	Dr. Pankaj Bhardwaj							
	Co-Principal Investigator (Co-PI)	Dr. Puneeta Pandey							
ix.	Project Implementing Partners								
	Key Persons / Point of Contacts with Contact Details, Ph. No, E-mail	Dr. Pankaj Bhardwaj Asst. Professor, Dept. of Botany, Central University of Punjab, Bathinda. Email: pankajihbt@gmail.com Phone No. 9501686709							

2. Project Outcome

2.1. Abstract (not more than 500 words) [it should include background of the study, aim, objectives, methodology, approach, results, conclusion and recommendations).

***Background*:** *Taxus wallichiana* and *Ulmus wallichiana* are endangered and vulnerable Himalayan species respectively, so require immediate conservation before they reach to the brink of extinction. Integration of Molecular and climatic data can provide better outcomes to get useful insights about the population and landscape genetics of these species, which a prior requirement for conservation planning. Further, Plants respond to environment and adapt accordingly by reshaping their gene pools as well as by bringing alterations in gene expression and its regulation, elucidation of which is beneficial.

Objectives/ Aim:

- To generate reference transcriptomes for *T. wallichiana* and *U. wallichiana*.
- To generate a polymorphic SSR marker resource.
- To utilize the polymorphic markers in elucidation of population and landscape genetics of these species.
- To reveal the effect and involvement of environmental factors in shaping the genetic diversity and population differentiation through geospatial technologies.

Methodology: We used RNAseq approach to construct de novo assemblies followed by unigenes generation. The unigenes were structurally and functionally annotated. In this way, a well annotated reference transcriptome was generated for both species. The immediate use of the reference transcriptome allowed us to screen and identify SSR regions. Primers were then designed for SSR regions. Following characterization of the SSRs, a set of polymorphic SSR markers was generated for both species. These polymorphic markers were used to elucidate the population and landscape genetic parameters and this molecular data was integrated with climatic and environmental data generated through ecological niche modelling. The integrative analysis yielded a better strategy for successful conservation of these species. Moreover, the adaptability of T. wallichiana to changing environment was elucidated through gene expressional study and taxonomic ambiguity of its family, Taxaceae was resolved through the use of reference transcriptome.

***Results*:** An exhaustive field survey was carried out to collect *Taxus wallichiana* and *Ulmus wallichiana* samples from the Indian Himalayas. SSR markers were generated and characterized for polymorphism for both the species. Further, the polymorphic SSR markers along with geospatial climate were utilized for elucidation of population and landscape genetic analysis to propose robust conservation strategies. The climatically suitable and genetically more diverse patches were identified which are best suited for conservation. Threat factors were revealed for *T. wallichiana*. In addition to this, we also elucidated the environmental impact on transcriptome flexibility of *T. wallichiana* for local adaptation under changed

climate through RNAseq based differential approach. Also, the taxonomic ambiguity of the family of Taxus wallichiana i.e. Taxaceae was phylogenetically resolved utilizing a more robust approach. The details of the work carried out, methodology, results and discussions are presented at length in the following sections.

Conclusion: Our study provided reference transcriptomes for both species which would open the door for further exploration of these species. We provided polymorphic SSR markers and strategy for effective conservation.

***Recommendations**:*

We recommend that the inferences generated here for conservation shall be followed for successful conservation.

2.2. Objective-wise Major Achievements

S. No.	Objectives	Major achievements (in bullets points)
1.	Sample collection and germplasm isolation	• More than 400 samples were collected from western Himalayas and 80 samples from eastern Himalayas. • Their DNA was isolated and stored. In this way a DNA bank of these two species was created.
2.	Micro-level and macro-level spatial mapping of *T. wallichiana* and *U. wallichiana*	• Spatial mapping of probability distribution of *T. wallichiana* was carried. • Maps depicting the suitable range for current and future distributions were generated.
3.	Change detection study to ascertain the changes in vegetation pattern.	• Change detection was carried out to observe the direction of change in probability of distribution from current to future (2070). • Maps depicting the patches with greater probability and patches with degraded probability of suitable distribution were generated.
4.	Transcriptome sequencing for *T. wallichiana* and *U. wallichiana*	• RNAseq based approach was used to generate the transcriptomic sequences for both species. • The raw reads were stored in the lab repository. • The quality of the reads was good.

5.	Assembly and annotation	• The raw reads were de novo assembled to create full length transcripts. • Following sequence redundancy removal and clustering, more than one lakh unigenes were generated for each species. • The unigenes were structurally and functionally annotated. • In this way a well annotated reference transcriptome was created for both the species. • These reference transcriptomes may provide basis for future research in exploring them at the genomic level.
6.	Prediction of micro-satellites, primer designing and characterization on populations.	• The reference transcriptomes were screened for SSR regions. • We identified more than 7000 and 14000 SSRs in *T. wallichiana* and *U. wallichiana* respectively. • Primers were designed using BatchPrimer3 for 100 and 90 randomly selected SSRs in *T. wallichiana* and *U. wallichiana* respectively. • Characterization of these SSRs was carried out on 30 and 20 samples representing 3 and 2 populations in *T. wallichiana* and *U. wallichiana* respectively through non-denaturating PAGE followed by silver staining. • Characterization resulted into generation of a set 37 and 28 polymorphic SSR markers in *T. wallichiana* and *U. wallichiana* respectively. • These SSRs have high potential for being utilized in population and landscape genetics of these species.

7.	Population and landscape genetic analysis of *T. wallichiana.*	• Among the 37 polymorphic SSRs of T. wallichiana, a set of 20 SSRs was utilized to reveal its population and landscape genetics in western Himalayas on 241 samples constituting 17 populations. • Moderate genetic diversity was revealed. • Occurrence of recent genetic bottleneck was observed. • Prevalence of inbreeding was found. • Observed heterozygosity was lower than expected. • Patches containing most of the alleles (genetic diversity hotspots) were identified.
8.	Ecological niche modelling in *T. wallichiana*.	• Modelling was carried out using Maxent model. • Distribution maps were generated for current and future time periods depicting the probability of climatic suitable for distribution of T. wallichiana. • Comparative analysis of the maps at the two time periods revealed the occurrence certain patches that are to remain stable or get improved (climatic hotspots) for *T. wallichiana* in future and certain patches that would degrade climatically so would be less favorable for distribution of *T. wallichiana.*
9.	Integrative analysis of genetic and climatic data to create an effective conservation strategy in *T. wallichiana*	• We provided a conservation strategy for rescuing this species through focusing our attention on genetic diversity hotspots and climatic hotspots. • Our strategy would be cost effective, save labour time and ensure conservation under safe environment.

10.		• The study to reveal the flexibility, switching and swing of the transcriptomic profile under temporally changing environment during different seasons for *T. wallichiana* was carried out. • A contrasting variation in expression of genes related to various biological processes and metabolic pathways enabled *T. wallichiana* to survive under changing environment. • The alternative splicing augments the gene expression in local adaptation and survival. • The oscillations within the transcriptome under natural conditions provided insights into the ecological adaptation of *T. wallichiana* as well as its performance.
11.	Resolving the taxonomic ambiguity of the family of *T. wallichiana*	• A large set of orthologous genes (331) were identified through comparatively transcriptomic analysis. • Alignment of these orthologs generated a sufficiently long gapless alignment for reconstructing the phylogeny of Taxaceae and Cephalotaxaceae. • Our study is more robust, has more power and higher statistical significance to ensure maintaining their familial distinctiveness.
12.	Elucidation of pattern of codon usage bias in *T. wallichiana*	• The analysis of codon usage bias revealed a tendency towards the preference of A/U ending in *T. wallichiana*. • Our study showed that natural selection is a predominating factor. • The mutational pressure, gene length, Gravy, Aromo and nucleotide composition are other contributing factors govern the codon usage bias weakly in *T. wallichiana*.

2.3. Outputs in terms of Quantifiable Deliverables*

S. No.	Quantifiable Deliverables*	Monitoring Indicators*	Quantified Output/ Outcome achieved	Deviations made, if any, and Reason thereof:
1.	DNA		The germ plasm of more than 300 samples belonging to both species was preserved.	
2.	RNA		RNA under different environmental stimuli is preserved	
3.	Reference transcriptomes		Well annotated reference transcriptomes are stored in lab repository	
4.	SSR markers		Polymorphic SSR markers are stored	
5.	Predicted maps		Environment suitability maps have generated.	

(*) As stated in the Sanction Letter issued by the NMHS-PMU.

2.4. Strategic Steps with respect to Outcomes (in bullets)

S. No.	Particulars	Number/ Brief Details	Remarks/ Enclosures
1.	New Methodology developed	A unique RNAseq based phylotranscriptomic pipeline was established to identify the gene orthology for phylogeny analysis.	
2.	New Models/ Process/ Strategy developed		
3.	New Species identified	Phylogenetic reconstruction of Taxaceae and Cephalotaxaceae lead to the conclusion that the family distinctiveness of both these families shall be maintained. The idea of merging these families into one, is not justified by our study.	
4.	New Database established	Transcriptomic sequences were generated and submitted to NCBI SRA database.	

S. No.	Particulars	Number/ Brief Details	Remarks/ Enclosures
5.	New Patent, if any		
	I. Filed (Indian/ International)		
	II. Granted (Indian/ International)		
	III. Technology Transfer(if any)		
6.	Others (if any)		

3. Technological Intervention

S. No.	Type of Intervention	Brief Narration on the interventions	Unit Details (No. of villagers benefited / Area Developed)
1.	Development and deployment of indigenous technology		
2.	Diffusion of High-end Technology in the region		
3.	Induction of New Technology in the region		
4.	Publication of Technological / Process Manuals		
	Others (if any)		

4. New Data Generated over the Baseline Data

S. No.	New Data Details	Status of Existing Baseline	Additionality and Utilisation New data
1.	RNAseq data for *T. wallichiana* and *U. wallichiana*	Well annotated reference transcriptomes was lacking for both these species. Our analysis bridged the gap.	
2.	SSR markers	While the polymorphic SSRs for T. wallichiana were very scarce, for U. wallichiana, they were completely lacking. Our generated markers are sufficient in number and informative to conduct diversity studies anywhere in the world.	

3.	Predicted maps	Climatically suitable maps depicting the change in environment, either favourable or degradative in future for both species was completely lacking. Our analysis provided guide to focus conservation on specifically required patches.	

5. Demonstrative Skill Development and Capacity Building/ Manpower Trained

S. No.	Type of Activities	Details with number	Activity Intended for	Participants/Trained			
				SC	**ST**	**Woman**	**Total**
1.	Workshops						
2.	On Field Trainings						
3.	Skill Development						
4.	Academic Supports	**2**	**PhD**				**2**
	Others (if any)						

6. Linkages with Regional & National Priorities (SDGs, INDC, etc)/ Collaborations

S. No.	Linkages /collaborations	Details	No. of Publications/ Events Held	Beneficiaries
1.	Sustainable Development Goal (SDG)			
2.	Climate Change/INDC targets			
3.	International Commitments			
4.	Bilateral engagements			
5.	National Policies			
6.	Others collaborations			

7. Project Stakeholders/ Beneficiaries and Impacts

S. No.	Stakeholders	Support Activities	Impacts
1.	Gram Panchayats		
2.	Govt Departments (Agriculture/ Forest)		
3.	Villagers		
4.	SC Community		
5.	ST Community		
6.	Women Group		

	Others (if any)		

8. Financial Summary (Cumulative)

S. No.	Financial Position/Budget Head	Funds Received	Expenditure/ Utilized	% of Total cost
I.	Salaries/Manpower cost	660000.00	494872.00	75%
II.	Travel	200000.00	187557.00	93%
III.	Expendables &Consumables	649115.00	692821.00	108%
IV.	Contingencies	60000.00	45618.00	76%
V.	Activities & Other Project cost	NIL		
VI.	Institutional Charges	NIL		
VII.	Equipments	900000.00	884977.00	98%
	Total	2469115.00		
	Interest earned	104332.00		
	Grand Total	2573447.00		

* Please attach the consolidated and audited Utilization Certificate (UC) and Year wise Statement of Expenditure (SE) separately, *ref*. **Annexure I.**

9. Major Equipment/ Peripherals Procured under the Project** (if any)

S. No.	Name of Equipments	Cost (INR)	Utilisation of the Equipment after project
1.	Centrifuge Machine	394147.00	Machine will be continue in use after the project
2.	Work Station	268105.00	Computing facility will be continued to support the other projects.
3.			
4.			
5.			

Details should be provided in details **(*ref* **Annexure III &IV).**

10. Quantification of Overall Project Progress

S. No.	Parameters	Total (Numeric)	Remarks/ Attachments/ Soft copies of documents
1.	IHR States Covered	3	Details Attached
2.	Project Site/ Field Stations Developed		
3.	New Methods/ Modeling Developed		
4.	No. of Trainings arranged		
5.	No of beneficiaries attended trainings		
6.	Scientific Manpower Developed (Phd/M.Sc./JRF/SRF/ RA):	9	Details Attached
7.	SC stakeholders benefited		
8.	ST stakeholders benefited		
9.	Women Empowered		
10.	No of Workshops Arranged along with level of participation		
11.	On field Demonstration Models initiated	.... (attach maps about location & photos)	
12.	Livelihood Options promoted		
13.	Technical/ Training Manuals prepared		
14.	Processing Units established	.... (attach photos)	
15.	No of Species Collected	2	Images attached
16.	New Species identified		
17.	New Database generated (Types):	2.	Accession numbers attached on a separate sheet
	Others (if any)		

11. Knowledge Products and Publications:

S. No.	Publication/ Knowledge Products	*Number*		*Total Impact Factor*	*Remarks/* Enclosures
		National	International		
1.	Journal Research Articles/ Special Issue:		2	12	attached
2.	Book Chapter(s)/ Books:				
3.	Technical Reports				
4.	Training Manual (Skill Development/ Capacity Building)				
5.	Papers presented in Conferences/Seminars				
6.	Policy Drafts/Papers				
7.	Others:		4		Under review

* Please append the list of KPs/ publications (with impact factor and further details) with due Acknowledgement to NMHS.

12. Recommendation on Utility of Project Findings, Replicability and Exit Strategy

Particulars	Recommendations
Utility of the Project Findings	See section 8.1
Replicability of Project	See section 8.2
Exit Strategy	Refer section 8 for details

(**PROJECT PROPONENT/ COORDINATOR**)

(Signed and Stamped)

(**HEAD OF THE INSTITUTION**)

(Signed and Stamped)

Place: ……..
Date: …../……/……

PART B: PROJECT DETAILED REPORT

The Detailed report should include an Executive Summary and it should have separate chapters on (i) Introduction (ii) Methodologies, Strategy and Approach (iii) Key Findings and Results (iv) Overall Achievements (v) Project's Impacts in IHR (vi) Exit Strategy and Sustainability (vii) References and (viii) Acknowledgement (It should have a mention of financial grant from the NMHS, MoEF&CC)

Further, description of Technical Activities, List of Trainings/ Workshops/ Seminars with details of trained resources, list of New Products developed under the project, Manual of Standard Operating Procedures (SOPs) developed, Technology developed/Transferred etc should be enclosed as Appendix.

2 EXECUTIVE SUMMARY

Taxus wallichiana, a famous anticancer drug synthesizing plant, is distributed along the Western Himalayas in the temperate zone. *Ulmus wallichiana* is an economically important vulnerable Himalayan species. Despite the immense industrial and medicinal value of *T. wallichiana*, and considerable economic importance of *U. wallichiana*, they are scarcely explored at the genomic and transcriptomic level. This study was focused to generate a structurally and functionally well annotated reference transcriptome for both species which would be useful in addressing different scientific problems for this species. In this study, using the RNA sequences several questions have been addressed.

Climate change, human over exploitation, deforestation, diseases and other environmental variables pushed *T. wallichiana* and *U. wallichiana* to their limits and it is at the verge of extinction. These factors severely reduced their populations which has profound impact on its gene pool, genetic diversity and genetic differentiation. To explore this, a set of highly polymorphic SSR markers was generated for them, by utilizing the reference transcriptome, for elucidation of its underlying population and landscape genetic parameters. This information on genetic diversity and structure is a prerequisite for planning the conservation strategies of this species. Further, identification of the genetic diversity hotspots was necessary to focus the conservation on these patches containing most of the alleles. Moreover, the climate change has been a big problem for the survival of the endangered species which prompted to identify the climatic hotspots which remain stable or get improved for *T. wallichiana*. Thus, an integrative approach of molecular and climatic analysis yielded a more suitable and better conservation strategy for *T. wallichiana*.

The survival under varied environment and climatic variables involves changes at the genetic and physiological networks for local adaptation. Apart from the exploring the effect of environment on genepool and genetic diversity of *T. wallichiana* spatially, the study also revealed the flexibility, switching and swing of the transcriptomic profile under temporally changing environment during different seasons. A contrasting variation in expression of genes related to various biological processes and metabolic pathways enabled *T. wallichiana* to survive under changing environment. The alternative splicing augments the gene expression in local adaptation and survival. The key challenge in the plant biology is to predict the adaptive responses taking place under natural environments. This study revealed the dynamics of the transcriptome occurring under natural conditions constituting a diverse array of

concurrent stimuli. Further, the oscillations within the transcriptome under natural conditions provided insights into the ecological adaptation of *T. wallichiana* as well as its performance.

The reference transcriptome generated in this study not only answers spatial and temporal impact effect of environment on genepool and expression profile respectively, but also has the potential for addressing other problems at the genomic level like taxonomy and analysis of codon usage bias. Since the taxonomic ambiguity of the family of *T. wallichiana* remains unresolved, this study undertook the advantage of power of RNAseq to address this problem by utilizing the reference transcriptome. A large set of orthologous genes were identified which generated a sufficiently long gapless alignment for reconstructing the phylogeny of Taxaceae and Cephalotaxaceae. Our study is more robust, has more power and higher statistical significance to ensure maintaining their familial distinctiveness. Moreover, the analysis of codon usage bias revealed a tendency towards the preference of A/U ending in *T. wallichiana*. Our study showed that natural selection is a predominating factor while mutational pressure, gene length, Gravy, Aromo and nucleotide composition are other contributing factors govern the codon usage bias weakly in *T. wallichiana*. Unequal utilization of synonymous codons is a well-known phenomenon among the living organisms. Enhancement in accuracy and efficiency of translation is a major role of this phenomenon. Understanding the degree and determining forces of codon usage bias in *Taxus contorta* will prove useful in interpreting the evolutionary characters of this species.

3 INTRODUCTION

3.1 Background of the Project (max. 500 words)

Taxus wallichiana and *Ulmus wallichiana* are endangered and vulnerable Himalayan species respectively. Overexploitation for drug extraction and other uses have rendered these species at the brink of extinction. If the present conditions prevail longer, these species may be wiped out form the globe. Immediate need id to turn our attention towards conservation of these species. In order to achieve this SSR marker designing through RNAseq approach can applied in elucidation of population and landscape genetic analysis of these species. Integration of Molecular and climatic data can provide better outcomes to get useful insights about the population and landscape genetics of these species, which a prior requirement for conservation planning. Moreover, there is very scarce data st the genomic and transcriptomic level of these species. They remain unexplored in this aspect. Use of RNAseq approach can be cost effective and beneficial to generate reference transcriptomes for these species which would suffice the gap to a greater extent in absence of reference genomes, if not completely. Further, in nature, plants seldom experience the controlled conditions, rather their response is triggered against the wholesome of the environment. The dynamics of the transcriptome under controlled conditions reflects only a small part of the entire display occurring under natural conditions. The assessment of oscillations within the transcriptome under natural conditions not only offers insights into the ecological adaptation of the plants but also performance. In view of this, an attempt to decipher the swing of the expressional profile in would ensure elucidation of better adaptive response under changing environment. Also, there is taxonomic ambiguity in the family of *T. wallichiana*. The current debate is whether Taxaceae and

Cephalotaxaceae shall be merged into a single family or retained into separate families. Use of the generated reference transcriptome can prove handy in the comparative transcriptomic analysis to resolve this issue.

3.2 Overview of the Major Issues to be Addressed (max. 1000 words)

1. **Transcriptome Characterization and Development of Functional Polymorphic SSR Marker Resource for Himalayan Endangered Species, *Taxus contorta* (Griff)**.

Taxus contorta is an important medicinal plant currently listed as endangered in IUCN Red Data List. It produces an anticancer drug, paclitaxel which is well known in the industrial sector. Due to habitat destruction and overexploitation, it is at the verge of extinction. Genomic and transcriptomic data for this species is scarce which has hampered its genomic studies. Moreover, large scale polymorphic informative codominant marker resource is also scarce which hinders its population and landscape genetic analysis. Here, we generated a reference transcriptome for this species which would facilitate the understanding of the functional elements and promote genomic research in this species. Also, a robust polymorphic SSR marker resource was characterized which can be used in conservation of this species. More than 100 million paired end raw reads were obtained through Illumina sequencing. A total of 129869 unigenes with mean sequence length of 1244nt were obtained from 209860 *denovo* assembled transcripts. Of these, 35752 transcripts were assigned 5971 unique GO terms. Around 40386 transcripts were found to have 2163 unique Pfam Ids. Pathway analysis against KEGG database yielded 3721 unique enzyme numbers. Screening of the transcripts for microsatellite regions yielded 7041 SSRs. Among the 100 SSRs selected for characterization on 30 genotypes, 37 polymorphic markers showed a total of 214 alleles with mean of 5.78 alleles per locus. Mean effective number of alleles (Ne) was found to be 3.64 and average PIC value of 0.64 was observed. Observed heterozygosity (0.57) was found to be lower than expected (0.69). This effective polymorphic SSR marker resource will act as valuable tool for deciphering its genetic diversity.

2. **Transcribing Molecular and Climatic data into conservation management for Himalayan endangered species, Taxus contorta.**

Owing to the changing climatic scenario globally and human overexploitation, the risk of extinction of Himalayan endangered species has amplified many folds. Taxus contorta, an endangered gymnosperm has reached to critical state in Western Himalayas, therefore needs immediate and robust management to rescue it. This study is aimed to elucidate the population and landscape genetics of T. wallichiana and planning its successful conservation strategies. Through an integrative approach of molecular genetics and climate ecology, complex interaction of genetic constitution and environment in T. wallichiana was elucidated. We used SSR markers to elucidate the genetic diversity hotspots and ecological niche modelling to elucidate the climatic hotspots of T. wallichiana from Indian Western Himalayas. The genetic

bottleneck, low heterozygosity and enhanced homozygosity were identified. We propose that the changing climate can cause failure of an entire conservation management programme, if the shift creates a degradative environment in future at the sites of conservation. Further, the conservation management is futile if it fails to enhance or maintain the genetic diversity. We propound that the use of germplasm from the genetic diversity hotspots for propagation in climatic hotspots and prioritization of these hotspot patches for conservation will ensure greater genetic variability under safe environment in the T. wallichiana during the conservation management. This integrative approach of translating molecular and climatic data into conservation planning will save our efforts, time and capital investment and ensures greater success in managing the revival of T. wallichiana in Western Himalayas.

3. Adaptive response of *T. wallichiana* under changed environment

In nature, plants seldom experience the controlled conditions, rather their response is triggered against the wholesome of the environment. The dynamics of the transcriptome under controlled conditions reflects only a small part of the entire display occurring under natural conditions. The assessment of oscillations within the transcriptome under natural conditions not only offers insights into the ecological adaptation of the plants but also performance. In view of this, here we attempted to decipher the swing of the expressional profile in *Taxus contorta* to ensure better adaptive response under changing environment during different seasons. The abundance estimation using RNAseq approach revealed 6727 differentially expressed genes. Comprehensive reprogramming was observed in taxol biosynthesis, maintainace of redox homeostasis and generation of effective shield to UV-B, high light intensity and temperature and. In addition to the differential expression, alternative splicing and single nucleotide variations (SNVs) also confer flexibility to the transcriptome of *T. wallichiana*. 1936 differentially expressing transcripts were also found to exhibit Differential Exon Usage (DEU) and SNVs. LC-MS based untargeted metabolic analysis revealed 7774 ion features among which around 334 putatively identified metabolites were differentially regulated. These oscillations provide flexibility to *T. wallichiana* for better survival under changed environment.

4. Analysis of codon usage bias in *T. wallichiana*

Unequal utilization of synonymous codons is a well-known phenomenon among the living organisms. Enhancement in accuracy and efficiency of translation is a major role of this phenomenon. Gymnosperms are rarely paid attention in this aspect. Understanding the degree and determining forces of codon usage bias in Taxus contorta, an endangered Himalayan gymnosperm will prove useful in interpreting the evolutionary characters of this species. Based on RNAseq data, from a total of 93790 assembled transcripts, 32701 unigenes were generated among which 13061 full length sequences were utilized for the analysis of codon usage bias. Mean ENC value revealed that codon usage bias is not strong in T. wallichiana. The preferred codons showed a trend of A/U ending while avoided codons showed G/C ending. Our study showed that natural selection is a predominating factor while mutational pressure,

gene length, Gravy, Aromo and nucleotide composition are other contributing factors govern the codon usage bias weakly.

5. RNAseq based phylogenetic reconstruction of Taxaceae and Cephalotaxaceae

Taxaceae and Cephalotaxaceae are the two economically important coniferale families. Over the years there has been much controversy over the issue of merging these families. Position of Amentotaxus and Torreya is also ambiguous. Some authors consider them closer to Taxaceae while others deemed them fit within Cephalotaxaceae. Still others prefer to raise them to their own tribe. Different morphological, anatomical, embryological and phylogenetic evidences support one or the other view, making the precise delineation between them unresolved. Here we used an RNAseq based approach to obtain orthologous genes across the selected species to reconstruct a more robust phylogeny of these families. A total of 233.123 million raw reads were denovo assembled to generate nine different transcript assemblies for the corresponding species. Out of the 940191 assembled transcripts across nine species, we generated 409734 unigenes which were clustered into orthologous groups. A total of 331 single copy complete orthologous groups were selected for phylogenetic analysis. Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian phylogenetic trees showed a sister relationship between Taxaceae and Cephalotaxaceae. Our analysis support their family distinctiveness and also shows that Amentotaxus and Torreya fits within Cephalotaxaceae.

6. Transcriptome Characterization and Development of Functional Polymorphic SSR Marker Resource for Himalayan Vulnerable Species, *Ulmus wallichiana*.

Ulmus wallichiana is a vulnerable species which requires attention to conserve it in advance before it reaches at the verge of extinction. Genomic and transcriptomic resources for this species are lacking which hinder the genetic exploration of this species. Further, no polymorphic marker resource is available for this species which can be used for elucidation of its underlying genetic diversity, required prior to conservation management of a threatened species. Here, we generated a reference transcriptome for this species to promote its genomic research and also utilized the reference transcriptome for the generation of a set of polymorphic SSR markers resource. Around 6.6 million paired end raw reads were de novo assembled into 605789 transcripts which were clustered into 146083 unigenes. Of these, 19909 transcripts were assigned 3986 unique KEGG ids, 70519 transcripts with 6621 unique Pfam domains, 45125 transcripts with 7302 unique INTERPRO domains and 1456 transcripts were identified as transcriptions factors (TFs) which belong to 53 unique TF families. Further, 8868 unique GO terms were obtained for the unigenes. A total of 16570 SSRs were identified. Out of 90 SSRs selected for characterization on 20 genotypes, 28 were polymorphic. Mean effective alleles (Ne) of 2.53, mean observed heterozygosity (Ho) of 0.77, mean Shannon index (I) of 0.95 and average polymorphic information content (PIC) of 0.57 were found. This study is the first report on marker resource development and transcriptomic data in U. wallichiana.

3.3 Baseline Data and Project Scope (max. 1000 words)
3.4 Project Objectives and Target Deliverables (as per the NMHS Sanction Order)

4 METHODOLOGIES, STARTEGY AND APPROACH

4.1 Methodologies used for the study

1. Sampling

Sampling of *Taxus* and *Ulmus w* from the Himalayas was carried out was carried out during the period 2016-2018. More than 500 samples were collected for both species. The details of the samples are provided in Table 1 and Table 2 in Appendix 1.

2. Transcriptome Characterization and Development of Functional Polymorphic SSR Marker Resource for Himalayan Endangered Species, *Taxus wallichiana* and *Ulmus wallichiana*.

2.1 RNA isolation and Sequencing

Samples for RNA isolation of Taxus and Ulmus were collected from the forest in liquid nitrogen and transported to the lab. RNA was isolated using the protocol described by Kejani et al, (2010) with some modifications. Quality and quantity of the isolated RNA was assessed through Qubit Flurometer and Nanodrop Spectrophotometer respectively. The integrity of the RNA was checked using Agilent Bioanalyzer chip. Leaf samples for DNA isolation were collected from different locations in Indian Himalayan regions (details of sampling points are given in Appendix 1). DNA was isolated through CTAB method developed by Doyle, (1991) with some modifications. The quality and quantity was assessed through Nanodrop spectrophotometer ND-1000 and agarose gel electrophoresis. The isolated DNA was stored at -20^{0}C for subsequent use.

2.2 Library prepration and sequencing

RNA sequencing libraries were prepared with Illumina-compatible NEBNext® Ultra™ Directional RNA Library Prep Kit (New England BioLabs, MA, USA). 1 µg of total RNA was taken for mRNA isolation, fragmentation and priming followed by synthesis of first and second strands. The double stranded cDNA was then ligated with Illumina Universal Primers as per NEBNext® Ultra™ Directional RNA Library Prep Kit protocol. The adaptor ligated fragments were enriched

to produce a sequencing library. The libraries from two samples were then sequenced on Illumina HiSeq 2000 platform.

2.3 Denovo assembly

Cleaning of raw reads was done by removing low quality bases and adaptors through Trim Galore version 0.4.1. FastQC was used to check the quality of the raw reads. Concatenation of the quality trimmed reads from individual samples was done prior to the construction of assembly. *Denovo* assembly from the concatenated and cleaned raw reads was constructed through Trinity version 1.6 (Grabherr *et al*., 2011). We used Read Representation, Blast against SwissProt and BUSCO for quality assessment of the assembly. Read representation was achieved by mapping the raw reads back to the assembly using Bowtie2 version2.3.0 (Langmead and Salzberg, 2012). Blast was done for counting the number of full length transcripts and BUSCO v2 (Simao *et al*., 2015) was used for assessing the completeness of the assembly. Removal of the redundant sequences and generation of unigenes was done through CD-HIT-EST version 4.6 (Li and Godzik, 2006) at 95% sequence identity threshold. The schematic representation of the pipeline used is shown in materials and methods in Appendix 1.

2.4 Functional annotation of unigenes

The non-redundant assembly was then functionally annotated using the annotation pipeline, Annocript (Musacchia et al, 2015). We used customised homology search against viridiplantae at an e-value of 0.00001. KEGG IDs were assigned to the transcripts by executing Blast through KAAS (Moriya et al, 2007). Further, the assignment of transcripts to gene families was done using the pipeline TRAPID (Van-Bel et al, 2013) with PLAZA2.5 as a reference database at an e-value of 10e-5. Protein domains were identified from Pfam database. For the identification of transcription factors, the transcripts were aligned against the PlantTFDB v4.0 (Jin et al, 2017) at an e-value of 0.00001.

2.5 SSR screening and characterization

The assembled transcripts were screened for microsatellite markers using MISA (Beier et al, 2017). Positional distribution of the SSRs in the transcripts was analysed by predicting the ORFs from the SSR containing sequences using orfPredictor (Min et al, 2005) followed by correlating the SSR start and end positions with the start and stop positions of the predicted ORFs using an in-house python script. The primers for the SSR containing sequences were designed using BatchPrimer3 (You et al, 2008). The characterization of the synthesized primers was performed on 30 genotypes, representing three populations, through non denaturing PAGE followed by silver staining using the protocol developed by Huang et al, (2018). Bands were scored and the raw data was analyzed for the calculation of marker parameters like PIC using Cervus (Marshall

et al, 1998), Observed and expected heterozygosity, Shannon's information Index and Diversity statistics using Popgene v3.2 (Yeh et al, 1999) and GenAlex v6.5 (Peakall and Smouse, 2012).

2.6 Cross species transferability

For cross species transferability analysis, SSRs were also screened through MISA from *T.baccata* genome assembly retrieved from oneKP (https://sites.google.com/a/ualberta.ca/onekp/) which is a consortium responsible for sequencing of over 1000 plants. The identified SSRs were then characterized on the selected genotypes of *T.wallichiana* using non denaturing PAGE.

3. Transcribing Molecular and Climatic data into conservation management for Himalayan endangered species, *Taxus wallichiana*.

3.1 Genetic diversity and population structure

The leaf samples of *T. wallichiana* were collected during the period 2016-2018 from the Indian Western Himalayan forests. The DNA of around 241 samples comprising 17 populations was subsequently isolated and used for further analysis. The details of sampling locations are presented in Appendix 1. Twenty developed in-house SSR markers were utilized for genotyping overall populations through polyacrylamide gel electrophoresis followed by silver staining (Huang et al. 2018). Primary population genetic parameters were determined by PopGene v3.2 (Yeh et al. 1999) and GenAlex v6.5 (Peakall & Smouse 2006, 2012). The spatial genetic structure and optimum number of genetic stocks were determined by STRUCTURE v2.3.4 (Pritchard *et al.*, 2000) and Structure Harvester (Earl, 2012) respectively. Genetic diversity hotspots were identified through interpolation of observed heterozygosity using Inverse Distance Matrix (IDM) tool implemented in ArcGIS 10.3.

3.2 Integrative analysis of genetic and climatic data

In order to assess the spatial relation of the bioclimatic variables and genetic diversity of *T. wallichiana* populations, we used the current bioclimatic variables retrieved from WorldClim version2 (Fick and Hijimans, 2017) at a spatial resolution of 30 seconds and analyzed their correlation with genetic diversity to reveal the most profound variables impacting the genetic diversity spatially. PCoA was used to cluster the populations based on bioclimatic variables to assess the isolation by climate theory (IBC). The climate based clustering was compared with genetic clustering to assess the relation of climate distinctiveness on genetic distinction. We used Mantel's test to find a possible link between genetic distance and geographic distance. The

isolation by distance may be due to the effect of isolation by climate because the relation between genetic and geographic distance could be due to selection by environment, since populations far away may experience distinct climate. The extent of correlation between the geographic distance and climatic distance was carried out to give us an idea about the proportion of impact by environment on isolation by distance.

3.3 Bottleneck analysis

Owing to the endangered status of *T. wallichiana* and its fragmented and isolated populations, it is necessary to perform demographic analysis to find any recent contraction in population as it has several consequences on the future of the populations. Several mutations model are applied for this purpose, however TPM is recommended for SSR markers over IAM or SMM. Msvar, Mratio and Bottleneck tests are the currently available methods for detecting genetic bottleneck in the populations. Here, we used all the three approaches to evaluate the past demographic changes in *T. wallichiana* in Indian Western Himalayas. For heterozygosity tests using Bottleneck, TMP model was used and run at 30% variance and 90% proportion of SMM for 1000 iterations. Mratio test was conducted with $\theta = 4$, *ps* = 90% and $\delta g = 3.5$. For Msvar, generation time was set to 5, mutation rate at 10^{-5} and the assumed value of time since the populations decline or expand was set to 10^5 years.

3.4 Ecological modelling

The GPS coordinates of *T. wallichiana* from West Himalayan regions were recorded through field visits during 2016-2018. The environmental variables were retrieved from WorldClim database. Predictors of the past (Midholocene), current and future environment were taken from WorldClim at a spatial resolution of 30seconds. Processing of the environmental predictor rasters was performed in R. We used Maxent (Phillips et al. 2006) model for predicting the suitability of the environmental predictors for *T. wallichiana* in Western Himalayas. Only 80% of the occurrence points were used as the training data to run the model. Rest of the 20 % was withheld as a test data. This was done to validate the accuracy of the predictions by the model. The model was evaluated by using the test data. An approach of area under receiver operator curve (AUC) was used to generate true positives and false positives. Successful prediction rate was generated by assessing true positives as a function of false positives. This generates a curve, AUC whose value ranges from 0 (no reliability) to 1 (perfect reliable). The change in suitability of the environment for the distribution of *T. wallichiana* in Western Himalayas was also assessed through comparison of predicted suitability at the above chosen temporal points.

4. Adaptive response of *T. wallichiana* under changed environment

4.1 RNA isolation, sequencing and denovo assembly

Leaf samples were collected from the forest in liquid nitrogen from Himachal Pradesh during early January and June thereby representing the two contrasting seasons, winter and summer respectively. RNA isolation, sequencing and denovo assembly was carried out in the same manner as described earlier.

4.2 Abundance estimation and identification of Differential expressed genes (DEGs)

The trinity assembled transcripts of all the samples were concatenated followed by removal of redundancy using CD-HIT-EST to create a master reference assembly for downstream analysis. Reads from the individual libraries were then aligned back to the reference transcriptome for count estimation of individual libraries using RSEM[1]with Bowtie2 as the alignment tool. A matrix of the read counts of individual libraries was generated which was subsequently analysed for differential gene expression using DESeq2 implemented in DEBrowser. First TMM based normalization was carried out on the estimated transcript counts. Transcripts with low expression (CPM < 1 in at least 1 samples) were filtered out. Batch effect correction was performed on filtered transcripts using Combat method. A fold change of ≥ 2 and an FDR adjusted *p* value of 0.001 was used for the estimation of differentially expressed genes (DEGs) under Wald test and TMM normalization.

4.3 Functional characterization of DEGs

GO annotations were transferred through TRAPID at an e-value of 10e-5. Visualization of these GO terms was performed through ReviGO. For visualization of individual genes associated with observed biological pathways in MapMan, we used Mercator4 to pass pathway functional annotation to our transcripts. The DEGs were also analysed for the identification of transcription factors through the sequence alignment with Plant TFDB.

4.4 Differential exon usage and SNP variation of DEGs

For differential exon usage analysis, we used an assembly first approach. The unigenes were used for mapping raw reads using Bowtie2. Next, Corset was used to generate clusters of the transcripts followed by generation of supertranscripts for each cluster using Lace v1.13. This supertranscriptome was then used to map the raw reads through STAR. FeatureCounts from Subreads R package was used to count the reads mapping to exonic regions and to generate a count matrix which was then subsequently used for differential exon usage analysis using DEXSeq. A thresh hold of 1% FDR (padj=0.01) was used to extract significant exonic regions.

For differential SNP analysis, we used mapping first approach to detect SNPs in our data using KisSplice which only outputs a local context around SNPs since it is not a full length assembler. The position of the identified SNPs was then obtained by aligning the SNPs detected by KisSplice with the Trinity assembled transcripts through BLAT. The unaligned SNPs were excluded from further analysis. To predict the functional impact of SNPs, KisSplice2RefTranscriptome was used. Detection of condition specific SNPs were obtained by KissDE.

4.5 Validation through quantitative Real Time PCR (qRT-PCR)

Using the High capacity cDNA reverse transcription kit (Applied Biosystems), cDNA was constructed from the isolated RNA according the manufacturer instructions. qRT-PCR primers were designed from the selected transcripts through Primer3Plus. We used PowerUp SYBER Green Master Mix (Applied Biosystems) for carrying out qRT-PCR of 11 transcripts with actin as a housekeeping gene. Fold changes were calculated using the $2^{-\Delta\Delta CT}$ method. Expression of these transcripts from RNAseq and qRT-PCR was then correlated to validate our results.

5. Analysis of codon usage bias in *T. wallichiana*

We used in-house sequenced Illumina based paired end RNAseq data (SRA accession SRX4951470) for conducting this study. Raw reads were trimmed for adaptor removal by using Trimmomatic v0.38 (Bolger et al. 2014) followed by generation of denovo assembly using Trinity v1.6 (Grabherr et al. 2011). The assembled transcripts were subjected to removal of sequence redundancy using CD-HIT-EST v4.6 at 95% sequence identity threshold (Li and Godzik 2006) and to generate unigenes. The full length transcripts of these non-redundant transcripts were identified through TRAPID (Van Bel et al. 2013) using plaza 2.5 database as reference in alignment. The coding sequences (CDS) of these full length transcripts containing a start and stop codon were then used for codon usage bias analysis.

The codon bias indices like RSCU and ENC, overall nucleotide composition and composition at the third codon position were determined by CodonWv1.4.2 and GCUA. The GC content at 1st, 2nd and 3rd positions of the codons was determined by using a Perl script (https://github.com/hxiang1019/calc_GC_content.git). ENC plot, neutrality plot and PR2 plot were used to determine whether the biasness is determined by mutation, selection or both. Correlations between ENC and GRAVY, ENC and Aromo and gene length and ENC were carried out to assess the impact of gene length and hydrophobicity on codon usage pattern. All the correlations were carried out in R.

6. RNAseq based phylogenetic reconstruction of Taxaceae and Cephalotaxaceae

We used an in-house sequenced Illumina based paired-end RNAseq raw reads of *Taxus wallichiana* together with the raw reads of eight species retrieved from NCBI SRA database for analysis. The schematic representation of the pipeline used for the orthology detection and phylogenetic analysis is shown in Appendix 1. The raw reads were cleaned by removing the adaptor sequences and low-quality bases through Trim Galore v0.4.1. The quality trimmed reads were then *denovo* assembled using Trinity v1.6 (Grabherr *et al*., 2011). The completeness of the assembly was analyzed using BUSCO v2 (Simao *et al*., 2015). Further, the raw reads were mapped back to the assembly using Bowtie2 v2.3.0 (Langmead and Salzberg, 2012) to assess the read representation for quality assessment. The quality was also checked by counting full-length transcripts through BLAST against SwissProt database. Removal of the redundant sequences and generation of unigenes was done through CD-HIT-EST v4.6 at 95% sequence identity threshold (Li and Godzik, 2006). The nonredundant assemblies were then compared with each other by executing reciprocal one to one blast at an e-value of 1e-05 for the identification of orthologous groups across the species through an orthology detection tool, Proteinortho v5.16b. All the degenerate or incomplete orthologous groups were filtered out to obtain only complete ones which contain no more than one gene from each species. The sequences of these single copy ortholog groups were processed to prepare the input files for alignment using in-house python scripts and then aligned individually by MAFFT v7.312. The individual aligned sequences were concatenated by AMAS v0.98 to generate a super-alignment file. Trimming of the spurious sequences and poorly aligned regions from the super-aligned sequences was performed using TrimAl v1.2. We used a multiple phylogenetic approach in the construction of phylogenetic tree. The maximum parsimony analysis was run on MEGA v7.0.26 using SPR search method. Support for the individual clades was achieved through bootstrapping with 500 replicates. Maximum Likelihood phylogenetic tree was constructed by IQ-TREE v1.5.5 with Ultrafast Bootstrap support (Minh *et al*, 2013; Haong *et al*, 2017) and SH-aLRT branch support for 10000 replicates. For model testing we used the ModelFinder implemented in the IQ-TREE. The Bayesian Markov-chain Monte Carlo (MCMC) analysis which consisted of two runs for 1000000 generations was performed by using Mr. Bayes v3.2.6. The default 25% burn-in was used to construct the tree under 50% majority-rule. The comparison and calculation of differences between the trees obtained from different methods was done by using TOPD/FMTS v3.3.

4.2 Preparatory Actions and Agencies Involved (max. 1000 words)

4.3 Details of Scientific data collected and Equipments Used (max 500 words)

The geographic coordinates of both *T. wallichiana* and *U. wallichiana* samples were recorded during field visits. Climatic data was retrieved from WorldClim for past, present and future at a spatial resolution of 30 seconds. RNAseq data yielded a primary data concerning the annotated assembled transcripts and SSRs. A myriad of equipments were used during the study like Horizontal and vertical electrophoretic systems, NanoDropspectrometer, GelDoc imaging system, PCR machines, Illumina HiSeq Sequencer platform, High end computation system, RT-PCR machine Water bath, Incubator, GPS, -80 and -20 refrigerator systems, Centrifuge machine, pH meter, Desiccator, etc.

4.4 Primary Data Collected (max 500 words)

The geographic coordinates of both *T. wallichiana* and *U. wallichiana* samples were recorded during field visits. Climatic data was retrieved from WorldClim for past present and future at a spatial resolution of 30 seconds. RNAseq data yielded a primary data concerning the annotated assembled transcripts and SSRs.

4.5 Details of Field Survey arranged (max 500 words)
4.6 Strategic Planning for each Activities (max. 1000 words)
4.7 Activity wise Time frame followed [using Gantt/ PERT Chart (max. 1000 words)]

5 KEY FINDINGS AND RESULTS

5.1 Major Research Findings (max. 1000 words)

1. **Transcriptome Characterization and Development of Functional Polymorphic SSR Marker Resource for Himalayan Endangered Species, *Taxus wallichiana* and *U. wallichiana***

High quality denovo assemblies comprising 129869 and 146083 unigenes for *T. wallichiana* and *U. wallichiana* were generated respectively. These unigenes were structurally and functionally annotated to create a reference transcriptome for both species. Screening of the unigenes for microsatellite containing loci showed that about 6507 sequences contain 7041 SSRs out of which 534 were in compound form and 800 sequences contained >1 SSR in *T. wallichiana*. Around 14042 sequences were found to contain 16570 SSRs out of which 1318 (7.95%) were in compound form and 2094 contained more than 1 SSRs in *U. wallichiana.* A total of 100 SSR markers for *T.wallichiana* were tested for characterization on 30 genotypes comprising three populations. Around 49 SSRs (61.25%) showed successful amplification

among which 27 markers showed polymorphism while 22 revealed monomorphism. 15 out of 20 markers designed from genomic sequences of *T.baccata* showed successful amplification across *T.wallichiana* genotypes with 10 as polymorphic and 5 as monomorphic. Thus a total of 75% cross transferability rate was achieved between the *T.baccata* and *T.wallichiana*. 90 primer pairs were selected randomly for genomic DNA amplification on 20 genotypes of *U. wallichiana*. Approximately 57 primer pairs showed successful amplification with 28 as polymorphic and 29 as monomorphic.

2. Transcribing molecular and climatic data into conservation management for Himalayan endangered species, Taxus contorta.

A set of 20 polymorphic SSR markers was utilized for elucidation of population and landscape genetic analysis of 241 samples comprising 17 populations of T. wallichiana from Western Himalayas. A moderate diversity with mean observed heterozygosity of 0.68 was obtained. While the mean observed heterozygosity was found to be slightly lower than expected, the mean observed homozygosity was higher than expected. A positive value (0.12) was found for mean fixation index (F). The mean Fst value of 0.19 indicates substantial genetic differentiation. AMOVA test revealed 14% of the variance contained among the populations and 86% within populations. Fst based genetic clustering revealed that populations segregated on the basis of geographic proximity. Similar results were revealed in Principal Coordinate Analysis (Appendix 1, Fig. 8A). The genetic stratification using multilocus data set through STRUCTURE utilizing a Bayesian model based approach using admixture model grouped individuals based on similar patterns of variation. Structure harvester revealed that there are two optimal genetic groups encompassing all the populations based on the highest value of delta K (Appendix 1, Fig. 9). Based on the membership coefficient arbitrary values, the pure ($Q \geq 80$ %) and admixed ($Q < 80$ %) percentage was found to be 65.6% and 34.4 % respectively. One interesting feature in Dendrogram and PCoA is that the population clusters and sub clusters are in line with geographic positions suggesting a direct link between population differentiation and geographic distance. This was revealed during Mantel test where a significant correlation value of 0.61 and p value of 0.01 was observed. The necessity of identifying the genetic diversity hotspots for conservation purposes is crucial. In this regard we identified three hotspots having high diversity (Appendix 1, orange to red, Fig. 10D) and three hotspots with substantial diversity (Appendix 1, yellow Fig. 10D). Interestingly, the genetic diversity in J&K populations is quite good as compared to HP and UK.

Analysis of reduction in population size was performed on individual local demes as well as on whole Himalayan range taken as single population. Our analysis revealed that, in the Western Himalayan range, *T. wallichiana* has undergone genetic bottleneck. All the tests performed confirm the population decline at the gross level in whole western Himalayas as a single population. Msvar analysis indicated that this decline had started around 50118 years ago. In the assessment of reduction in size in individual local demes, Mratio test and Msvar analysis revealed bottleneck in 16 and 17 populations respectively.

Analysis of individual demes showed that the reduction in individual populations became significant around 1148 to 6918 years ago.

The assessment of the predicted change in suitability of the environment showed that at different temporal points, certain areas show degradation in suitability while others reveal an improvement. We observed that compared to the current environment governing the spatial distribution of *T. wallichiana*, there occurs degradation at many points while other points show improvement of the environmental predictors in both past and future.

Our study throws light on the underlying genetic diversity pattern of *T. wallichiana* and its associated climatic niche which govern its distribution. Few concerns identified in this study which alarm the incoming threat to this important species were identified here;

- Enhanced homozygosity and reduced heterozygosity due to inbreeding.
- Extremely fragmented nature of populations found during the field visits. In most of the forests during our field visits we found only few standing individuals present in a relatively large forest area. Much of the population has disappeared.
- Genetic bottleneck
- Degradation of favorable environment in large number of patches in future for *T. wallichiana*.

The above mentioned key concerns appeal for immediate and robust conservation of this medicinally, industrially and ecologically important endangered Himalayan gymnosperm. Based on our analysis, it is recommended that:

- The genetic diversity hotspots identified shall be focused for preservation in Indian Western Himalayan region. This will ensure preservation of most of the alleles owing to their greater genetic diversity.
- Other moderate and low diversity areas need attention for conservation but require restoration of genetic diversity first through propagation of genetically diverse germplasm collected from the hotspot areas.
- The areas showing stability or improvement in context of climate in future shall be prioritized. Restoration of *T.wallichiana* populations in these patches shall be immediate focus of the local authorities.
- The patches which are likely to get degraded environmentally, shall not be prime focus of conservation management to save efforts, time and capital. However, the allelic diversity in these patches shall be preserved either in gene banks or by raising such genotypes in stable areas.
- There is a tendency towards inbreeding. Consequently, the nurseries propagating the Taxus either by means of seeds or vegetatively through stem cuttings should be examined for genetic diversity base hosted in them and subsequently renewed for promoting genetic diversity.

3. Adaptive response of *T. wallichiana* under changed environment

Assessment of variation in expression profiles under changed environment in *T. wallichiana* showed a total of 6727 genes were significantly differentially expressed at fold change of ≥ 2 and a *padj* value of 0.001. Among them, 2805 genes were upregulated while 3921 genes were found to be downregulated in winter as compared to summer. Functional characterization revealed that among the important processes that are upregulated in summer include response to UV-B and high light intensity, transition from vegetative to reproductive phase etc. The processes involved in the production of secondary metabolites like MVA, MEP, Shikimate, and Phenylpropanoid pathways etc are also found to function better in summer than in winter. In the downregulated category, there were 75 transcripts corresponding to 31 TF families while 51 transcripts corresponding to 21 TF families were found in upregulated category. Besides, prominent reprogramming to confer better adaptability under prevailing conditions occurs in context of compatible solutes, response to UV-B, changes in cell wall, response to cold, response to photoperiods, reorganization of the membrane components etc.

Taxol biosynthesis pathway is specifically confined to the members of Taxaceae. We found the transcripts corresponding to this biosynthetic pathway and its precursors, all down regulated in winter. Further, the transcripts corresponding to the JA pathway which incites taxol biosynthetic pathway are found upregulated in summer. Moreover, *i*n summer, the response to necrotrophic pathogens is observed to be enhanced as the transcripts corresponding to PAMP and DAMP were found upregulated.

Around 20518 exons corresponding to 34973 transcripts (28.35%) were differentially used under summer and winter. 3747 transcripts also have differential expression. A total of 21334 SNVs corresponding to 17645 transcripts (14.30%) were found to be significantly differential between summer and winter at a p-value of 0.001 and a minimum Δf (frequency difference of SNPs between conditions) of 0.1%. Around 8681 transcripts have SNVs within CDS region. 4283 SNPs were non-synonymous producing 373 unique codon changes. 2509 transcripts with differential exon usage were also observed to have differential expression among which 743 contain non-synonymous SNPs. Thus a set of 1936 transcripts have both differential expression and differential exon usage as well as condition specific SNP variation

4. Analysis of codon usage bias in *T. wallichiana*

Based on RNAseq data, from a total of 93790 assembled transcripts, 32701 unigenes were generated among which 13061 full length sequences were utilized for the analysis of codon usage bias. Mean ENC value revealed that codon usage bias is not strong in *T. wallichiana*. The preferred codons showed a trend of A/U ending while avoided codons showed G/C ending. Our study showed that natural selection is a predominating factor while mutational pressure, gene length, Gravy, Aromo and nucleotide composition are other contributing factors govern the codon usage bias weakly.

5. RNAseq based phylogenetic reconstruction of Taxaceae and Cephalotaxaceae

RNAseq based approach to obtain orthologous genes across the selected species was used to reconstruct a more robust phylogeny of Taxaceae and Cephalotaxaceae. A total of 233.123 million raw reads were denovo assembled to generate nine different transcript assemblies for the corresponding species. Out of the 940191 assembled transcripts across nine species, we generated 409734 unigenes which were clustered into orthologous groups. A total of 331 single copy complete orthologous groups were selected for phylogenetic analysis. Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian phylogenetic trees showed a sister relationship between Taxaceae and Cephalotaxaceae. Our analysis support their family distinctiveness and also shows that Amentotaxus and Torreya fits within Cephalotaxaceae.

5.2 Key Results (max 1000 words in bullets covering all activities)

7. We generated a reference transcriptome for *T. Wallichiana* and *U. wallichiana*. The reference transcriptomes would open the door for exploration of the species in genomic field. This transcriptomic resource can prove useful in further genetic analysis of these species. The transcriptomes were structurally and functionally well annotated. Over one lakh unigenes were generated for each species.
8. The reference transcriptomes were first utilized for the screening and discovery of SSR markers. More than 7000 and 16000 SSR markers were identified for *T. Wallichiana* and *U. wallichiana* respectively. This is the first marker resource for *U. wallichiana*. Characterization of the identified markers for polymorphic information yielded a set of 37 and 28 SSR polymorphic markers for *T. Wallichiana* and *U. wallichiana* respectively. These markers are very informative and have the potential to be employed in population and landscape genetic analysis of these species.
9. A set of 20 polymorphic SSR markers was used for elucidation of diversity patterns of *T. wallichiana* in Western Himalayas on 241 samples comprising 17 populations. A moderate level of genetic diversity was observed. Observed heterozygosity was found to be less than expected which could be attributed to inbreeding. Patches containing most genetic diversity (genetic diversity hotspots) were identified. A recent genetic bottleneck was also observed for the populations of *T. Wallichiana*. Climatic analysis through ecological niche modelling yielded predicted suitability maps which reveal the patches that are going to be stable or get improved (Climatic hotspots) for *T. wallichiana* in future along Western Himalayas. Finally, integrating the genetic and climatic data resulted into a strategy that can prove comparatively successful in conserving this species saving time, capital, labour under safe environment.
10. Adaptive responses to environmental changes induces manipulation at the expressional level. Comprehensive reprogramming was observed in taxol biosynthesis, maintainace of redox homeostasis and generation of effective shield to UV-B, high light intensity and temperature and.

In addition to the differential expression, alternative splicing and single nucleotide variations (SNVs) also confer flexibility to the transcriptome of *T. wallichiana*. 1936 differentially expressing transcripts were also found to exhibit Differential Exon Usage (DEU) and SNVs. These oscillations provide flexibility to *T. wallichiana* for better survival under changed environment.

11. Based on RNAseq data, codon usage bias analysis in *T. wallichiana* revealed that codon usage bias is not strong in *T. wallichiana*. The preferred codons showed a trend of A/U ending while avoided codons showed G/C ending. Our study showed that natural selection is a predominating factor while mutational pressure, gene length, Gravy, Aromo and nucleotide composition are other contributing factors govern the codon usage bias weakly.

12. RNAseq based approach was used to obtain orthologous genes across the selected species to reconstruct a more robust phylogeny of these Taxaceae and Cephalotaxaceae. A total of 233.123 million raw reads were denovo assembled to generate nine different transcript assemblies for the corresponding species. Out of the 940191 assembled transcripts across nine species, we generated 409734 unigenes which were clustered into orthologous groups. A total of 331 single copy complete orthologous groups were selected for phylogenetic analysis. Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian phylogenetic trees showed a sister relationship between Taxaceae and Cephalotaxaceae. Our analysis support their family distinctiveness and also shows that Amentotaxus and Torreya fits within Cephalotaxaceae.

5.3 Conclusion of the study (maximum 500 words in bullets)

- Our study enabled to widen the field of exploration of *T. wallichiana* and *U. wallichiana* by providing functionally annotated reference transcriptomes for them. This can prove useful deciphering novel genetic elements and underlying diversity in both species.
- Further, our study provided a robust polymorphic SSR marker resource for both species which to the best of our knowledge is the first to develop such large set of polymorphic loci in both species. Our strategy for conservation of *T. wallichiana* is more reliable owing to the fact that it takes into account bot genetic diversity as well as climatic change. We believe that this strategy would be cost effective, save time and labour and is safe in the context of climate change.
- Elucidation of transcriptome flexibility under changed environment has enabled us to understand the key insights about its adaptive response. Moreover, the analysis of codon usage bias revealed natural selection as dominant factor determining the CUB.
- Further, our unique approach of orthology detection and their utility in phylogeny enabled us to resolve the taxonomic ambiguity of Taxaceae and Cephalotaxaceae.

6 OVERALL ACHIEVEMENTS

6.1 Achievement on Project Objectives [Defining contribution of deliverables in overall Mission (max. 1000 words)]

The achievements of study are as follows:

- Successful generation reference transcriptomes for *T. wallichiana* and *U. wallichiana*. The reference transcriptomes were structurally and functionally annotated. The details of these references transcriptomes are presented in in Appendix 1 under results section.
- Successful generation of polymorphic SSR markers for both species the details of which are presented in Appendix 1 under results section.
- Developing a comparatively better strategy for conservation of *T. wallichiana*. The details of which are presented in Appendix 1 under results section.
- Determining dominant factors in governing the codon usage bias in *T. wallichiana*. Details are presented in Appendix 1 under results section.
- Taxonomic ambiguity of Taxaceae and Cephalotaxaceae was resolved due to which their family distinctiveness shall be maintained. The details are presented in Appendix 1 under results section.
- Adaptive response of *T. wallichiana* under changed environment was elucidated. Details are presented in Appendix 1 under results section.

6.2 Establishing New Database/Appending new data over the Baseline Data (max. 1500 words, in bullet points)

We established transcriptomic data for *T. wallichiana* and *U. wallichiana*. More than one lakh unigenes were generated for each species. These unigenes were functionally and structurally annotated. This is the first transcriptomic sequence data generated for *U. wallichiana*. For *T. wallichiana*, this represents the most reliable and credible annotated RNAseq resource. Data for SSR regions was generated, their polymorphism is also revealed.

6.3 Generating Model Predictions for different variables (if any) (max 1000 words in bullets0
6.4 Technological Intervention (max 1000 words)
6.5 On field Demonstration and Value-addition of Products (max. 1000 words, in bullet points)
6.6 Promoting Entrepreneurship in IHR
6.7 Developing Green Skills in IHR
6.8 Addressing Cross-cutting Issues (max. 500 words, in bullet points)

7 PROJECT'S IMPACTS IN IHR

7.1 Socio-Economic Development (max. 500 words, in bullet points)
7.2 Scientific Management of Natural Resources In IHR (max. 500 words, in bullet points)

7.3 Conservation of Biodiversity in IHR (max. 500 words, in bullet points)

- Our study identified genetic diversity hotspots which contain most of alleles of the populations.
- Climatically stable or improved patches were identified.
- Our strategy for conservation holds that focus of the conservation programme shall be on the hotspots which would preserve most of the genetic diversity under safe environment.
- Other low diversity areas shall be addressed by propagating the germplasm from hotspot areas for promoting genetic diversity.
- The genotypes at climatically degraded areas shall be preserved either in gene banks or their multiplication in hotspots areas.

7.4 Protection of Environment (max. 500 words, in bullet points)

- Our study is focused on the protection of the threatened species, *T. wallichiana* and *U. wallichiana* which form and integral biotic component of Western Himalayan environment. Molecular and climatic analysis was used to provide an optimal rescue plan for them. In this way a strategy for protecting this biotic component was framed.

7.5 Developing Mountain Infrastructures (max. 500 words, in bullet points)
7.6 Strengthening Networking in IHR (max. 700 words, in bullet points)

8 EXIT STRATEGY AND SUSTAINABILITY

8.1 How effectively the project findings could be utilized for the sustainable development of IHR (max. 1000 words)

The key finding of our study in the context of sustainable development is the development of a comparatively better strategy for rescuing *T. wa*llichiana from extinction. Our strategy would ensure the continuous and longer supply of the species to the humanity. Thus, human society can be benefited from the wide industrial and medicinal properties of this species in a continue and sustainable manner. Our strategy not only takes into account the preservation of this species, but also maintaining its genetic diversity which is highly recommended for endangered species which otherwise cannot withstand environment change. Further, change in climate is also taken into account for preserving this species. Thus, this study is useful for sustainable development of this species.

8.2 Efficient ways to replicate the outcomes of the project in other parts of IHR (Max 1000 words)

Our methods are easily replicable. The SSR markers developed in this study can be employed to elucidate the genetic diversity of the samples collected from other areas of IHR. For this, samples need to collected, DNA isolated and finally using the developed set of polymorphic SSR markers for revealing the genetic diversity and structure. Additionally, a new set of markers can be characterized for polymorphic information by designing the primers for SSR regions of the reference transcriptome. After revealing the genetic parameters, the climatic data of the respective areas needs to be at hand to draw conservation inferences by integrative analysis of molecular and climatic data.

8.3 Identify other important areas not covered under this study needs further attention (max 1000 words)

Eastern Himalayan regions are other regions yet to be covered. Our approach can be applied to these regions for providing conservation planning.

8.4 Major recommendations for sustaining the outcome of the projects in future (500 words in bullets)

Few concerns identified in this study which alarm the incoming threat to this important species were identified here;

- Enhanced homozygosity and reduced heterozygosity due to inbreeding.
- Extremely fragmented nature of populations found during the field visits. In most of the forests during our field visits we found only few standing individuals present in a relatively large forest area. Much of the population has disappeared.
- Genetic bottleneck
- Degradation of favorable environment in large number of patches in future for *T. wallichiana*.

The above mentioned key concerns appeal for immediate and robust conservation of this medicinally, industrially and ecologically important endangered Himalayan gymnosperm. Based on our analysis, it is recommended that:

- The genetic diversity hotspots identified shall be focused for preservation in Indian Western Himalayan region. This will ensure preservation of most of the alleles owing to their greater genetic diversity.
- Other moderate and low diversity areas need attention for conservation but require restoration of genetic diversity first through propagation of genetically diverse germplasm collected from the hotspot areas.

- The areas showing stability or improvement in context of climate in future shall be prioritized. Restoration of *T.contorta* populations in these patches shall be immediate focus of the local authorities.
- The patches which are likely to get degraded environmentally, shall not be prime focus of conservation management to save efforts, time and capital. However, the allelic diversity in these patches shall be preserved either in gene banks or by raising such genotypes in stable areas.
- There is a tendency towards inbreeding. Consequently, the nurseries propagating the Taxus either by means of seeds or vegetatively through stem cuttings should be examined for genetic diversity base hosted in them and subsequently renewed for promoting genetic diversity.

9 REFERENCES/BIBLIOGRAPHY

Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics*, *33*(16), 2583-2585.

Doyle, J. (1991). DNA protocols for plants. In Molecular techniques in taxonomy (pp. 283-293). Springer, Berlin, Heidelberg.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, *4*(2), 359-361.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, *37*(12), 4302-4315.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I. & Chen, Z. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*. 29(7), 644.

Huang, L., Deng, X., Li, R., Xia, Y., Bai, G., Siddique, K. H., & Guo, P. (2018). A fast silver staining protocol enabling simple and efficient detection of SSR markers using a non-denaturing polyacrylamide gel. *Journal of Visualized Experiments*, (134), e57192.

Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., & Gao, G. (2016). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic acids research*, gkw982.

Kejani, A. A., Tafreshi, S. A. H., Nekouei, S. M. K., & Mofid, M. R. (2010). Efficient isolation of high quality nucleic acids from different tissues of *Taxus baccata* L. *Molecular biology reports*, *37*(2), 797.

Langmead, B., & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods, 9*(4), 357.

Li, W., & Godzik, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 22(13), 1658-1659.

Marshall, T. C., Slate, J. B. K. E., Kruuk, L. E. B., & Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular ecology*, *7*(5), 639-655.

Min, X. J., Butler, G., Storms, R., & Tsang, A. (2005). OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic acids research*, *33*(suppl_2), W677-W680.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic acids research*, *35*(suppl_2), W182-W185.

Musacchia, F., Basu, S., Petrosino, G., Salvemini, M., & Sanges, R. (2015). Annocript: a flexible pipeline for the annotation of transcriptomes able to identify putative long noncoding RNAs. *Bioinformatics*, *31*(13), 2199-2201.

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, 28, 2537e2539.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210- 3212.

Van Bel, M., Proost, S., Van Neste, C., Deforce, D., Van de Peer, Y., & Vandepoele, K. (2013). TRAPID: an efficient online tool for the functional and comparative analysis of de novo RNA-Seq transcriptomes. *Genome biology*, *14*(12), R134.

Yeh, F. C., Yang, R. C., Boyle, T., Ye, Z. H., & Mao, J. X. (1999). POPGENE, version 1.32: the user friendly software for population genetic analysis. *Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, AB, Canada*.

You, F. M., Huo, N., Gu, Y. Q., Luo, M. C., Ma, Y., Hane, D. & Anderson, O. D. (2008). BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC bioinformatics*, *9*(1), 253.

10 ACKNOWLEDGEMENT

We sincerely acknowledge MoEF & CC for providing the necessary grant for this study. We owe our thanks to Central University of Punjab, Bathinda for providing the necessary infrastructure during the study.

APPENDICES

Appendix 1 – Details of Technical Activities
Appendix 2 – Copies of Publications duly Acknowledging the Grant/ Fund Support of NMHS
Appendix 3 – List of Trainings/ Workshops/ Seminars with details of trained resources and dissemination material and Proceedings
Appendix 4 – List of New Products (utilizing the local produce like NTFPs, wild edibles, bamboo, etc.)
Appendix 5 – Copies of the Manual of Standard Operating Procedures (SOPs) developed
Appendix 6 – Details of Technology Developed/ Patents filled
Appendix 7 – Any other (specify)

******

Annexure-I

Consolidated and Audited
Utilization Certificate (UC) and Statement of Expenditure (SE)

For the Period: .....

1.	Title of the project/Scheme/Programme:	
2.	Name of the Principle Investigator & Organization:	
3.	NMHS-PMU, G.B. Pant National Institute of Himalayan Environment& Sustainable Development, Kosi-Katarmal, Almora, Uttarakhand Letter No. and Sanction Date of the Project:	
4.	Amount received from NMHS-PMU, G.B. Pant National Institute of Himalayan Environment& Sustainable Development, Kosi-Katarmal, Almora, Uttarakhand during the project period (Please give number and dates of Sanction Letter showing the amount paid):	
5.	Total amount that was available for expenditure (Including commitments) incurred during the project period:	
6.	Actual expenditure (excluding commitments) incurred during the project period:	
7.	Unspent Balance amount refunded, if any (Please give details of Cheque no. etc.):	
8.	Balance amount available at the end of the project:	
9.	Balance Amount:	
10.	Accrued bank Interest:	

Certified that the expenditure of **Rs._____ (Rupees _____)** mentioned against Sr. No. 6 was actually incurred on the project/scheme for the purpose it was sanctioned.

Date:

(Signature of Principal Investigator)	(Signature of Registrar/ Finance Officer)	(Signature of Head of the Organization)

OUR REF. No.

ACCEPTED AND COUNTERSIGNED

Date:

COMPETENT AUTHORITY
NATIONAL MISSION ON HIMALAYAN STUDIES (GBPNIHESD)

Statement of Consolidated Expenditure

[Institution Name here]

Statement showing the expenditure of the period from
Sanction No. and Date :

1. Total outlay of the project :

2. Date of Start of the Project :

3. Duration :

4. Date of Completion :

a) Amount received during the project period :
b) Total amount available for Expenditure :

S. No.	Budget head	Amount received	Expenditure	Amount Balance/ excess expenditure
1	Salaries			
2	Permanent Equipment Purchased (Item-wise			
3				
4				
5				
6				
7				
8				
9				
10	Institutional charges			
11	Accrued bank Interest			
12	**Total**			

Certified that the expenditure of **Rs._____ (Rupees:_____)** mentioned against Sr. No.12 was actually incurred on the project/ scheme for the purpose it was sanctioned.

Date:

(Signature of Principal Investigator)	(Signature of Registrar/ Finance Officer)	(Signature of Head of the Organization)

OUR REF. No.

ACCEPTED AND COUNTERSIGNED

Date:

COMPETENT AUTHORITY
NATIONAL MISSION ON HIMALYAN STUDIES (GBPNIHESD)

Annexure-II

Consolidated Interest Earned Certificate

Please provide the detailed interest earned certificate on the letterhead of the grantee/ Institution and duly signed.

Interest earned on Project Fund

Sr No	Date	Amount
1	03.06.2016	2700
2	02.09.2016	17204
3	02.12.2016	15945
4	02.03.2017	15070
5	02.06.2017	12580
6	01.09.2017	4881
7	02.12.2017	2986
8	02.03.2018	7589
9	03.06.2018	7132
10	02.09.2018	6126
11	02.12.2018	5121
12	04.03.2019	3949
13	04.06.2019	3049
	Total Interest	**104332**

(Signature of Principal Investigator)

(Signature of Registrar/ Finance Officer)

Annexure-III

Consolidated Assets Certificate

Assets Acquired Wholly/ Substantially out of Government Grants

10.1 (Register to be maintained by Grantee Institution)

Name of the Sanctioning Authority: _____

1. Sl. No. _____
2. Name of Grantee Institution: _____
3. No. & Date of sanction order: _____
4. Amount of the Sanctioned Grant: _____
5. Brief Purpose of the Grant: _____
6. Whether any condition regarding the right of ownership of Govt. in the property or other assets acquired out of the grant was incorporated in the grant-in-aid Sanction Order: _____
7. Particulars of assets actually credited _____ or acquired_____
8. Value of the assets as on _____
9. Purpose for which utilised at present _____
10. Encumbered or not _____
11. Reasons, if encumbered _____
12. Disposed of or not _____
13. Reasons and authority, if any, for disposal_____
14. Amount realised on disposal _____

Any Other Remarks: _____

_____

(PROJECT INVESTIGATOR) **(FINANCE OFFICER)**

(Signed and Stamped) **(Signed and Stamped)**

(HEAD OF THE INSTITUTE)

(Signed and Stamped)

Annexure-IV

List or Inventory of Assets/ Equipment/ Peripherals

S. No.	Name of Equipment	Quantity	Sanctioned Cost	Actual Purchased Cost	Purchase Details
1.	Centrifuge Machine	01	900000.00	394147.00	17.02.2017
2.	Work Station	01		268105.00	28.06.2017
3.	Satellite Data			222675.00	
			Total Rs	884977.00	

(PROJECT INVESTIGATOR)

(Signed and Stamped)

(FINANCE OFFICER)

(Signed and Stamped)

(HEAD OF THE INSTITUTE)

(Signed and Stamped)

Annexure-V

National Mission on Himalayan Studies (NMHS)

DIRECT BENEFIT TRANSFER (DBT) DETAILS

Scheme Name:	**National Mission on Himalayan Studies (NMHS)**
Scheme Type:	**Central Sector (CS) Grant-in-Aid Scheme**
Scheme Code:	**NMHS**
Category:	**Fellowship under Project Grant**
Month-Year:	

PRO FORMA FOR DBT DETAILS

University/Institution Name: .....

S#	Position (H-RA, H-JRF/ H-JPF)	Name	DoB*	DoI*	PI	Research title	Objectives	Study Area, IHR State	Contact details (Complete corresponding address), Mobile No., E-mail ID	Bank details (Account number, IFSC Code)	Emoluments /Fellowship	Aadhaar No.
1.												

Note: For each month, the DBT Details Pro forma dully filled and signed for each Himalayan Fellowship Grant under NMHS must be submitted at finance.nmhspmu2017@gmail.com; nmhspmu2016@gmail.com. *DoB (Date of Birth); DoJ (Date of Joining).

(Authorized Signatory)

Month **2019 – Latest Updated List of Himalayan Researchers or Fellows** *(working in the current time)*

S#	Name	Fellowship (RA/JRF/JPF)
1.		
2.		

Annexure-VIII

10.2

Technology Transfer and/ or Intellectual Property Rights Certificate

With a view to encourage the institutions to file patent applications on their innovations, motivate them to transfer their technologies for commercialization, and facilitate them to reward their inventions, the following instructions are issued.

1. In these instructions:

(a) **"Institution"** means any technical, scientific or academic establishment where research work is carried out through funding by the Central / State Government.

(b) **"Intellectual Property Rights"** include patents, registered designs, copyrights and layout design of integrated circuits.

(c) **"Inventor"** means an employee of the institution whose duties involve carrying out of scientific or technical research.

2. Scope: These instructions apply to those institutions receiving funds for research projects from the DBT, Ministry of Science and Technology.

3. Inventions by institutions: Institutions shall be encouraged to seek protection of Intellectual Property Rights (IPR) to the results of research through R&D projects. While the patent may be taken in the name(s) of inventor(s), the institutions shall ensure that the patent is assigned to it & DBT, GOI. The institution shall take necessary steps for commercial exploitation of the patent on non-exclusive basis. The institution is permitted to retain the benefits and earnings arising out of the IPR. However, the institution may determine the share of the inventor(s) and other persons from such actual earnings. Such share(s) shall be limited to 1/3rd of the actual earnings.

4. Inventions by institutions and industrial concerns: IPR generated through joint research by institution(s) and industrial concern(s) through joint efforts can be owned jointly by them as may be mutually agreed to by them and accepted by the Department through a written agreement. The institution and industrial concern may transfer the technology to a third party for commercialization on exclusive/non-exclusive basis. The third party, exclusively licensed to market the innovation in India, must manufacture the product in India. The joint owners may share the benefits and earnings arising out of commercial exploitation of the IPR. The institution may determine the share of the inventor(s) and other persons from such actual earnings. Such share(s) shall not exceed 1/3rd of the actual earnings.

5. Patent Facilitating Fund: The institution shall set apart not less than 25 per cent of such earnings for crediting into a fund called Patent Facilitating Fund. This Fund shall be utilized by the institution for updating the innovation, for filing new patent applications, protecting their rights against infringements, for creating awareness and building competency on IPR and related issues.

6. Information: The institutions shall submit information relating to the details of the patents obtained, the benefits and earnings arising out of IPR and the turnover of the products periodically to the Department/Ministry, which has provided funds.

7. Royalty-free license: The Government shall have a royalty-free license for the use of the intellectual property for the purposes of the Government of India.

(HEAD OF THE INSTITUTE)

(Signed and Stamped)