The first draft genome of *Picrorhiza kurrooa*, an endangered medicinal herb from Himalaya

Tanvi Sharma[1,3†], Nitesh Kumar Sharma[2,3†],Prakash Kumar[2,3], Ganesh Panzade[2,3], Tanuja Rana Guleria[1], Mohit Kumar Swarnkar[1], Anil Kumar Singh[4], Dharam Singh[1,3], Ravi Shankar[2,3]*, Sanjay Kumar[1,3]*

Affiliations:

[1]Biotechnology division,

CSIR-Institute of Himalayan Bioresource Technology (CSIR-IHBT),

Palampur (HP), 176061, India.

[2] Studio of Computational Biology & Bioinformatics,

CSIR-Institute of Himalayan Bioresource Technology (CSIR-IHBT),

Palampur (HP), 176061, India.

[3]Academy of Scientific and Innovative Research (AcSIR),

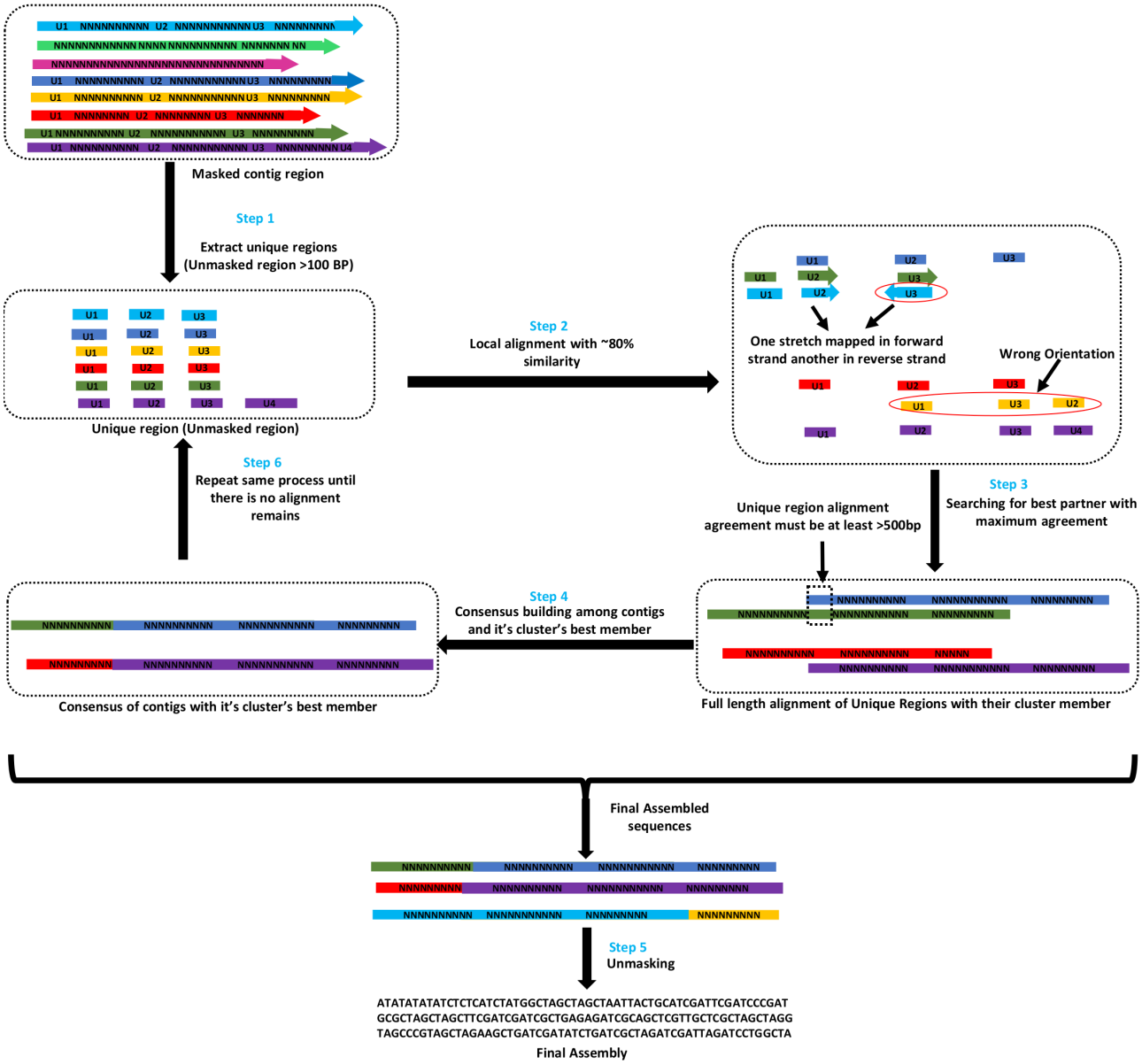
Ghaziabad, Uttar Pradesh- 201 002

[4]ICAR-Indian Institute of Agricultural Biotechnology, Ranchi

[†]Contributed Equally

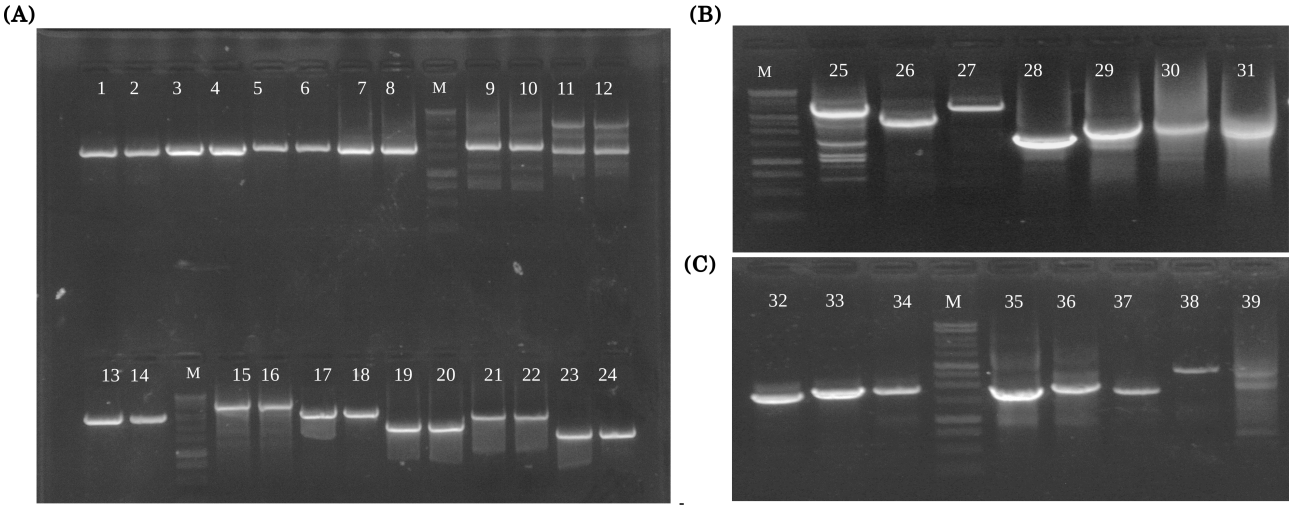
***Corresponding Author:** sanjaykumar@ihbt.res.in , ravish@ihbt.res.in

Supplementary files



Supplementary Figure 1 Schematic diagram of novel strategy used to join contigs into scaffold using unique stretches of contigs. In first step contigs unique regions were separated from repeats region, in second step these unique regions were aligned with each other using BLASTN with 80% similarity. In this step, alignments were checked for three criteria (i) orientation of alignment must be same in all unique stretches of one pair of contigs (ii) order of occurrence of unique region must be same in each pair of contig (iii) length of alignment must be >500 bp to avoid the occurrence of a fragment by chance. In third step best pairs for each contigs were selected. In forth step each were

merged into single contig, if best pairs having more than two contigs than consensus were built. This process repeated sixteen times and then in fifth step all scaffolds were unmasked.



Supplementary Figure 2 Gel photographs showing amplicons obtained from 3' and 5' ends of respective contigs (A,B,C). The well numbers in the gel correspond to superscript on amplicon size in Supplementary Table 2

Supplementary Table 1 Distribution of top 10 different species which were found most abundant from transcript BLAST tophit

Name of species	Number of transcript found in a species(%)
Sesamum indicum	8901 (36.31%)
Handroanthus impetiginosus	4447 (18.14%)
Erythranthe guttata	2186 (8.92%)
Olea europaea var. sylvestris	1184 (4.83%)
Dorcoceras hygrometricum	767 (3.13%)
Beta vulgaris subsp. vulgaris	534 (2.18%)

Ipomoea nil	435 (1.77%)
Chenopodium quinoa	274 (1.12%)
Helianthus annuus	126 (0.51%)
Cauliflower mosaic virus	119 (0.49%)

Supplementary Table 2 Table showing contig numbers, expected amplicon sizes, contig end amplified and standardized annealing temperature. The superscript on amplicon size corresponds to the well number shown in the gel in Supplementary figure 2

S. No	Contig Number	Expected PCR amplicon size in base pairs (bp)	Contig end amplified (5'end/3'end)	Annealing temperature (Ta in °C) with Advantage GC2 Polymerase (Takara)
1.	Contig_3840	1706 bp[1,2]	5'end	60
2.	Contig_5857	1876 bp[3,4]	3'end	60
3.	Contig_8777	1776 bp[5,6]	5'end	60
4.	Contig_2896	1812 bp[7,8]	5'end	60
5.	Contig_6284	2002 bp[9,10]	5'end	60
6.	Contig_6284	1811 bp[11,12]	3'end	60
7.	Contig_12645	2815 bp[13,14]	5'end	60
8.	Contig_115	4745 bp[15,16]	5'end	60
		3278 bp[17,18]	5'end	60
10.	Contig_4012	2734 bp[19,20]	5'end	60
		2210 bp[21,22]	3'end	55
12.	Contig_115	1599 bp[23,24]	3'end	52
13.	Contig_12645	2913 bp[25]	3'end	68
14.	Contig_5857	2824 bp[26]	5'end	60
15.	Contig_115	4666 bp[27]	5'end	55
16.	Contig_9572	1533 bp[28]	3'end	52
17.	Contig_8777	2102 bp[29]	3'end	60
18.	Contig_9572	2196 bp[30]	5'end	60
19.	Contig_3241	1882 bp[31]	5'end	60
20.	Contig_8754	1576 bp[32]	3'end	55
		1748 bp[33, 37]	5'end	55
22.	Contig_4012	1874 bp[34, 36]	5'end	55
23.	Contig_3241	1556 bp[35]	3'end	55
24.	Contig_2896	2760 bp[38]	5'end	55
25.	Contig_2896	2445 bp[39]	3'end	60

Supplementary Table 3 Primer sequences and PCR thermal cycling profile used for amplification of 3' and 5' end of contigs

Primer Name	Sequence 5'-3'	PCR thermal cycling profile
Contig_115Fwd1	ATTCTATAAAGTAAGGTAAGGATGATT	

Contig_115Rev2	AGTGATTTCAAGCCCTCTGCG	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min
Contig_115Fwd2	CGCAGAGGGCTTGAAATCACT	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 1 min 20 sec, 30 cycles; 72°C for 5min.
Contig_115Rev3	CTTTCTTTCCTTAACGACATCTCA	
Contig_115Fwd3	TGAGATGTCGTTAAGGAAAGAAAG	94°C for 4 min; 94°C for 30 sec, 55°C for 30 sec, 72°C for 1 min, 30 cycles; 72°C for 5min.
Contig_115Rev4	CTACGAAAAAAAAAAAAAAGAGAAAAGAGA	
Contig_115Fwd5	CGAAGAGCATTCACTTTCCATTG	94°C for 4 min; 94°C for 30 sec, 52°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_115Rev6	TGTGGTTGGGAACATGGACC	
Contig_12645Fwd1	CCATAGCTCGTTTCCCATAATAC	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_12645Rev2	TTGCACAAGTTTCTTCACAATTCC	
Contig_12645fwd3	AGCGTGGCTCGAACAGGC	94°C for 4 min; 94°C for 30 sec, 68°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_12645Rev4	TAGTGACGGAAATATCCGTCAC	
Contig_4012Fwd1	ATTCTTTAGGCTTGGAGCCCTT	

Contig_4012Rev2	AGACATATACAAATAGTTTATCGGTAG	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_4012Fwd2	CTACCGATAAACTATTTGTATATGTCT	94°C for 4 min; 94°C for 30 sec, 55°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_4012Rev3	TTCTTAAGGAACACATACTCACAC	
Contig_8777Fwd1	GTTTTCCTTTTTTTTACAGTGAGACC	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_8777Rev2	GGGTGATAGCATTGATAAAGGAAT	
Contig_8777Fwd3	CTCATACTAGAGTTGGAAGCTG	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_8777Rev4	CCAAACGACAATCCCAAACAGAA	
Contig_5857Fwd1	GTTAGAGAGGAAAGGATTCCTAG	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_5857Rev2	ACTGTGATTTTCATGCTTCCGCA	
Contig_5857Fwd 3	GTGGCTCCTTTGATTGATTCGT	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_5857Rev4	GAATCAAAGTATGACGAGTGCTC	
Contig_2896Fwd1	GATCTAGTGTTTCTCCAGGTTTC	94°C for 4 min;

		94°C for 30 sec, 55°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_2896Rev2	GAGTTGCATTGAAGTTCGCATC	
Contig_2896Fwd3	GATACGAGCAGTCATCAAGTCT	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_2896Rev4	ACCTAGCTAAGAGAGTGACTG	
Contig_2896Fwd5	AAGTCTCTCATATCTGCGCCAA	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_2896Rev6	CAGACAGACCTCAGAGAGATATT	
Contig_9572Fwd1	TTGATTCCCCTCTCATTCACTAC	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_9572Rev2	AGATATGTCCAATAGTTTTGAATGTTTGA	
Contig_9572fwd3	ATGAACTGCTGACAAAATAGAAATAAAA	94°C for 4 min; 94°C for 30 sec, 52°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_9572Rev4	AGAGTTGTTAAGTCTCGAGAGGAA	
Contig_3241Fwd1	GCCTTCCAGTGTTCATTTTCAG	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_3241Rev2	CAAGTACTTTGTGACTTTCATTGATG	

Contig_3241Fwd3	AATTTCATTTTTTCGAGTTTTACATCATG	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_3241Rev4	CATGATGTAAAACTCGAAAAAATGAAATT	
Contig_6284 Fwd1	GATGCTGCTGCTGAAAAGGC	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_6284 Rev2	TAAATATGCGTGTATCAAAAGTGTGT	
Contig_6284 Fwd3	TGCCTGATGCGATCTCCTGT	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_6284 Rev4	CCCTTAAATTTGAAGACCCAAGAA	
Contig_8754 Fwd1	CGTGCGTACGACACATTTGG	94°C for 4 min; 94°C for 30 sec, 55°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_8754 Rev2	ACGGTTCACACACTTTACTATTCA	
Contig_8754 Fwd4	TTCCAATTACCTTATCACCTGAATG	94°C for 4 min; 94°C for 30 sec, 55°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min
Contig_8754 Rev5	CAAGAGCTATATTCTTCTCCAATTTGATA	
Contig_3840 Fwd1	GACCCTATTCTTATCCGAAGTG	94°C for 4 min; 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min, 30 cycles; 72°C for 5min.
Contig_3840 Rev2	TCTTCAATGCCCACATCCATGA	

Supplementary Table 4 Primer sequences used for expression analysis of miRNA targets and miRNAs

Primers used for expression analysis of miRNA targets		
Prmer Name	**Sequence 5'-3'**	**PCR thermal cycling profile**
pkr043992 Fwd	CGGGAATTACAACGCAAAGTTATC	Initial denaturation at 94° C for 3 min, followed by 30 cycles of 94° C, 30 sec; 58° C, 30 sec; 72° C, 2 min. Final extension at 72° C for 5 min
pkr043992 Rev	TGGGCATGTTATATGGGATTCTC	
pkr030564 Fwd	GCCCGTGGATTTTCGAAA	Initial denaturation at 94° C for 3 min, followed by 30 cycles of 94° C, 30 sec; 58° C, 30 sec; 72° C, 2 min. Final extension at 72° C for 5 min
pkr030564 Rev	TTTCTTTGGCGCGATTAGGAT	
pkr028622 Fwd	CGAGAATTGGCAGAAGCAACTCG	Initial denaturation at 94° C for 3 min, followed by 30 cycles of 94° C, 30 sec; 58° C, 30 sec; 72° C, 2 min. Final extension at 72° C for 5 min
pkr028622 Rev	AAGTGGTTCGAGGCAGCAGT	
pkr009602 Fwd	GGTGTGGAGCGGGTGGAG	Initial denaturation at 94° C for 3 min, followed by 30 cycles of 94° C, 30 sec; 58° C, 30 sec; 72° C, 2 min. Final extension at 72° C for 5 min
pkr009602 Rev	AGAAGCTGCCCGGTGGTC	
pkr046121 Fwd	CCGCCACACATGAACATCAG	Initial denaturation at 94° C for 3 min, followed by 30 cycles of 94° C, 30 sec; 58° C, 30 sec; 72° C, 2 min. Final extension at 72° C for 5 min
pkr046121 Rev	CCGCGAGGAAGGAAAACTG	
pkr025510 Fwd	GGCCATTTGAGAAACCGATTT	Initial denaturation at 94° C for 3 min, followed by 30 cycles of 94° C, 30 sec; 58° C, 30 sec; 72° C, 2 min. Final extension at 72° C for 5 min
pkr025510 Rev	GATTGATCACCATTCCGTCAACT	
pkr043564 Fwd	CCAGTCGCCGCTACAACAC	Initial denaturation at 94° C for 3 min, followed by 30 cycles of 94° C, 30 sec; 58° C, 30 sec; 72° C, 2 min. Final extension at 72° C for 5 min
pkr043564 Rev	CGAACACCTGAACTTCCAAATAGA	

pkr051041 Fwd	GTTAAGAGCTTGTCGCCGAAA	Initial denaturation at 94° C for 3 min, followed by 30 cycles of 94° C, 30 sec; 58° C, 30 sec; 72° C, 2 min. Final extension at 72° C for 5 min
pkr051041 Rev	CGTCTCACAAAACCGTTGAAAA	
pkr043355 Fwd	CGACGTCATCCTCCTTTCAGA	Initial denaturation at 94° C for 3 min, followed by 30 cycles of 94° C, 30 sec; 58° C, 30 sec; 72° C, 2 min. Final extension at 72° C for 5 min
pkr043355 Rev	GGCGACGCCCTTTCTTTC	
pkActin Fwd	GGCTGGAAGAGCACCTCAGGG	94° C for 3 min, followed by 30 cycles of 94° C, 30 sec; 59° C, 30 sec; 72° C, 2 min. Final extension at 72° C for 5 min
pkActin Rev	CGTCTAAGACCAACTCGGCCGTC	
Primers used for expression analysis of miRNAs		
miRNA	**Sequence**	**Taqman Assay ID (Applied Biosystems**
pkr-miR1162-3p.2	AGTTGTAGGCTGTTGAAGAAGATC	CTAAAFC
pkr-miR-7385a-5p.1	TGGTCGGACTGCCTGAGGTCAGTC	CTTZ9CD
pkr-nov-miR-687.1	TTACATAGGATTGGCACGAAGC	CTMFWZN
pkr-miR-548ad-3p.2	AAAAACGAGAACCGAACACCGAAC	CTPRJ6H
pkr-miR4241.2	ATTTGGGAATGACGATTTAACACT	CTXGPZA
pkr-miR-5877.1	ATCGTATCCGAACTTTGTTGGCCC	CTZTD6Y
pkr-miR4241.1	ATTTGGGAATAACGCTTTAACACT	CT2W7RW
pkr-miR-378a-3p.1	ACTGGACTGGGTCTGAGTCGGACC	CT322CU
pkr-miR-578.1	ATTCTTGTGTTAGACTGTTTAACG	CT7DPGN
pkr-miR4241.6	ATTTGGGAATGACGTTTTAACACT	CT9HH2K
pkr 5.8s rRNA	CAACGGATATCTCGGCTCTC	CTCE3Y9